REMARKS

The Office is authorized to charge the fee for a three month extension of time to Deposit Account No. 50-0911. Any fees that may be due in connection with the filing of this paper or with this application may be charged to Deposit Account No. 50-0911. If a Petition for Extension of time is needed, this paper is to be considered such Petition.

A supplemental Information Disclosure Statement is filed under separate cover. References to which this response refers and that were not previously provided are provided herewith

Claims 1, 3-7, 9, 11-16, 45, 48, 51-53, 57-59, 61-63, and 65-78 are pending. Claims 1, 3-6, 16, 45, 51, 52, 53, 57, 59 and 63 are amended. The amendments are made for clarity to render it clear that the method identifies mutein proteases that are mutated so that their cleavage specificity is changed and it is changed to recognize a substrate sequence in a target protein, such that cleavage of **that sequence** in the target protein inactivates the target protein. As discussed below, the cited art does not teach or suggest modifying the specificity of a protease so that it cleaves a non-native cleavage sequence that is present in a target and that such cleavage inactivates the target. No new matter is added.

The arguments and response filed August 7, 2008, June 11, 2009, July 20, 2010 and April 13, 2011 are each incorporated by reference herein.

I. CLAIM OBJECTIONS

Claim 51 is objected to because it depends upon cancelled claim 50. This objection is addressed by amendment of the claims herein.

II. THE REJECTION OF CLAIMS 1, 3-7, 9, 11-16, 45, 48,, 51-53, 57-59, 61-63 and 65-78 UNDER 35 U.S.C. §112, FIRST PARAGRAPH – NEW MATTER

Claims 1, 3-7, 9, 11-16, 45, 48, 51-53, 57-59, 61-63 and 65-78 are rejected under 35 U.S.C. §112, first paragraph, because it is alleged that the claimed subject matter is not described in the specification. Specifically, the Examiner states that in Claim 1, step (a) with the different amino acid positions being mutated are not supported in the as-filed specification.

Applicant respectfully submits that the amendment to claims 1, 53, 59 and 63 to recite specific amino acid residues for mutation is supported in the specification as filed. For example, basis can be found at page 19, lines 27-29, which states:

> To make a variant protease with an altered substrate recognition profile, the amino acids in the three-dimensional structure that contribute to the substrate selectivity (**specificity determinants**) are targeted for mutagenesis. [emphasis added]

And at page 27, line 32 to page 28, line 3, which states:

> Essential amino acids in the proteases generated using the methods of the present invention are identified according to procedures known in the art, such as site-directed mutagenesis or saturation mutagenesis of active site residues. In the latter technique, residues that form the **S1-S4** pockets that have been shown to be important **determinants of specificity are mutated** to every possible amino acid, either alone or in combination. [emphasis added]

Further, basis can be found at page 25, lines 4-6, which states that S1-S4 amino acid residue specificity determinants can be mutated. The specification describes and identifies specificity determinants in serine proteases. For example, at page 19, line 33 to page 20, line 6, the specification recites:

> Structural determinants for various proteases are listed in Table 3, along with a listing of the amino acid of family members determined to be of known, extended specificity. For serine proteases, the following amino acids in the primary sequence are determinants of specificity: 195, 102, 57 (the catalytic triad); 189, 190, 191, 192, and 226 (P1); 57, the loop between 58 and 64, and 99 (P2); 192, 217, 218 (P3), the loop between Cys168 and Cys180, 215 and 97 to 100 (P4).

Table 3 at page 21 further sets forth specificity determinants as follows: S4 determinants include amino acid residue 171, 174, 180, 215, Cys165 to Cys182; S3 determinants include amino acid residue 192 and 218; S2 determinants include amino acid residues 99, 57 and 60s loop; and S1 determinants include amino acid residue 189, 190, 226 and Cys191 to Cys220.

Therefore, the specification as originally filed describes amino acid residues targeted for mutagenesis as claimed. Accordingly, reconsideration of this rejection respectfully is requested.

III. THE REJECTION OF CLAIMS 1, 3-7, 9, 11-16, 45, 48, 51-53, 57-59, 61-63 and 65-78 UNDER 35 U.S.C. §112, SECOND PARAGRAPH

Claims

Claims 1, 3-7, 9, 11-16, 45, 48, 51-53, 57-59, 61-63 and 65-78 are rejected under 35 U.S.C. §112, second paragraph as being indefinite for failing to particularly point out and distinctly claim the subject matter. Reconsideration of the grounds for this rejection is respectfully requested in view of the amendments herein and the following remarks.

Relevant law

Claims are not read in a vacuum but instead are considered in light of the specification and the general understanding of the skilled artisan. *Rosemount Inc. v. Beckman Instruments, Inc.*, 727 F.2d 1540, 1547, 221 USPQ 1, 7 (Fed. Cir. 1984), *Caterpillar Tractor Co. v. Berco, S.P.A.*, 714 F.2d 1110, 1116, 219 USPQ 185, 188 (Fed. Cir. 1983). Claim language is satisfactory if it reasonably apprises those of skill in the art of the bounds of the claimed invention and is as precise as the subject m*att*er permits. *Shatterproof Glass Corp. v. Libby-Owens Ford Col.*, 758 F.2d 613, 624, 225 USPQ 634, 641 (Fed. Cir.), cert. dismissed, 106 S.Ct. 340 (1985).

Claims 1, 53, 59 and 63

a. The Examiner states that the terns "each member of the library has N mutations relative to a wild-type mammalian protease scaffold or a biologically active portion thereof; and N is a positive integer," and "each different mutein protease in the library is a member of the library" are vague and indefinite. As suggested by the Examiner, these terms are deleted from the claim because they are redundant.

b. The Examiner states that clause (b) is indefinite as to whether the target is a protein or a polypeptide. Applicant respectfully submits that antecedent basis is clear. The target is a protein. Thus, clause (b) recites that the members of the library can be contacted with the "target protein" itself or with a portion of the target protein that is "a polypeptide comprising the substrate sequence that is present in the target protein." For example, clause (b) in claim 1 recites (in part):

> contacting members of the library with a target protein **or** with a polypeptide comprising a substrate sequence that is present in the target protein

c. The Examiner states that the term in clause (b) "the target protein is selected from among" is vague and indefinite because there "seems to be no selection that can be made as the members are 'among' those recited therein." Similarly, the Examiner states that the phrase "among amino acid residues between" is vague as to which residues are considered as between.

Applicant respectfully requests reconsideration of this rejection in view of the amendments to the claims herein. For example, in phrase b) the recitation of the target protein is amended to clearly recite that the listings are alternative classes of proteins or are alternative specific proteins. One of skill in the art is familiar with such classes of proteins and can identify a protein that meets one or more of the noted classes.

Further, with respect to the recitation "between" in clause a), the claim is amended for clarity to render it clear that "between" is only with reference to amino acid residues *between* amino acid 58 and amino acid 64. Also, the term "inclusive" is added for clarity to render it clear that selection of residues to mutate also can include residue 58 and 64, by chymotrypsin numbering, as well as intervening residues.

d. The Examiner states that the term "substrate sequence" is indefinite because it lack antecedent basis with the preceding step. Applicant respectfully requests reconsideration of this rejection in view of the amendment of the claims herein to mirror the language of clause (b) in clause (c) to recite that cleavage activity is measured for "the polypeptide comprising the substrate sequence that is present in the target protein."

e. The Examiner states that the term in step (d) of "based on the measured activity" is not positive, manipulative process. Applicant respectfully disagrees when step (d) is considered as a whole and with reference to step (c). In step (c), cleavage activity is measured, which is an affirmative step. In step (d), proteases are identified that have an increased activity and/or specificity (also an affirmative step), which is determined "based on the measured activity and/or specificity as determined in step (c). Nevertheless, for clarity, the claim is amended herein in order to render this relationship clear.

f. The Examiner states that clause (e) is indefinite because there is not nexus of this step with the preceding steps (a)-(d). The Examiner states that it is vague and indefinite as to the steps included and/or precluded in the testing.

Applicant respectfully disagrees. Step (d) is clearly linked to the preceding steps. For example, in step (d) proteases are identified that exhibit cleavage activity and/or specificity for the target protein or polypeptide containing a substrate sequence in the target protein, and in step (e) the identified proteases are tested for cleavage and inactivation of an activity to select for proteases that inactivate an activity of the target protein by the cleavage event. Nevertheless, the claim is amended herein to render this relationship more clear. Further, it respectfully is submitted that any method can be used to test for inactivation of an activity of the target protein, and thus the claim does not "include and/or preclude" any particular method. For example, as discussed in the previous response of record mailed April 13, 2011, numerous methods and assays are described in the specification and known to one of skill in the art to test for inactivation of an activity of a protein.

g. The Examiner states that the term "involved" is vague and indefinite with reference to a disease or pathology. It respectfully is submitted that based on the description in the specification, one of skill in the art would understand that "involved" in a disease or

pathology means that it contributes to the pathogenesis of a disease or pathology. Thus, it is a target protein that is a target for molecular intervention in order to treat the disease or pathology. For example, the specification at page 9, lines 23-27 recites:

> If a protease is engineered to recognize a substrate sequence within a target protein or proteins that would (i.) alter the function i.e. by inactivation of the target protein(s) upon catalysis of peptide bond hydrolysis and, (ii.) the target protein(s) are recognized or unrecognized as points of molecular intervention for a particular disease or diseases, then the engineered protease has a therapeutic effect via a proteolysis-mediated inactivation event.

Nevertheless, in the interest of advancing prosecution in this case, the claims are amended herein to recite that a target protein "involved in a disease or pathology" is one "whereby inactivating an activity of the target protein can treat the disease or pathology. Applicant respectfully submits that this is a limitation on the particular target protein or substrate sequence used in the method, and is not an affirmative step of the method. There is no limitation in the claim of any step of treating a disease or pathology.

Claims 3-6

The Examiner states that claims 3-6 are indefinite because the base claim does not recite the location, if any, of the N in the protease mutein. The claims are amended herein to recite the "number of mutation(s)" contained in the protease mutein in order to correct antecedent basis in view of the amendment of claim 1 herein. Further, it respectfully is submitted that it is clear to one of skill in the art that a protease that "comprises mutation(s)" from among any of the recited amino acid residues can contain a mutation at one of the amino acid residues, two, three, four, several, etc. or all of the amino acid residues. Claims 3-6 recite the number of mutations that are contained. Hence, the claims are clear.

Claim 16

The Examiner states that claim 16 is vague and indefinite as to step (e) given that the mutations are already recited in the base claim.

For clarity, claim 16 is amended to recite the further steps as steps (f) – (h), since claim 1 includes steps (a)-(e). In addition, claim 16 is amended to refer to step (d), which is the step in which protease muteins are identified in the method of claim 1. In view of these amendments, Applicant respectfully submits that the claim is clear. The protease muteins that are identified in step (d) by practice of the method are those members of the library that exhibit increased cleavage and/or specificity, and contain a mutation at one or more of the recited positions. Hence, step (d) is a selection/screening step to select for only those protease muteins in the library that exhibit increased activity/specificity, whereby the

mutation at the position(s) is likely linked to the observed increase in activity and/or specificity. Claim 16 is directed to identifying the *specific mutations* (position and amino acid mutation) contained in the identified members, and combining the specific mutations to generate a new mutein protease.

Claim 52

The Examiner states that claim 52 is indefinite for the term "in vivo assay," since the base claim does not recite for "in vivo" testing. Thus, the Examiner states that the claim is indefinite as to the metes and bounds of the in vivo assay.

Claim 52 is amended herein to render it clear that the testing for inactivation of an activity in an *in vivo* assay is a limitation on step (e), whereby testing the identified protease for inactivation of an activity of the target protein is in an in vivo assay. It respectfully is submitted that "testing" as recited in claim 1 can include any method or methods, *in vitro* or *in vivo*, in which a cleavage of a target protein can be assessed and/or activity of a target protein can be assessed and determined to be destroyed or inactivated when incubated in the presence of an identified protease.

IV. THE REJECTION OF CLAIMS 1, 3-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63 and 65-78 UNDER 35 U.S.C. §102

Claims 1, 3-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63 and 65-78 are rejected under 35 U.S.C. §102(b) as being anticipated by Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999)(as evidenced by Shi *et al.*, USP 20020197701). The Examiner states that Lien *et al.* discloses at pages 73-75 a method of identifying serine proteases using targeted combinatorial mutagenesis of serine proteinases by 1) producing a library of mutant enzymes; 2) contacting the library with a substrate; and 3) identifying the mutant. The Examiner states that Lien *et al.* discloses quantitative assessments of cleavage by monitoring hydrolysis of synthetic peptides, and also discloses that the mutant enzyme is useful for therapy as in blood coagulation. In this respect, the Examiner states that hydrolysis is analogous to "inactivation" as claimed (see bottom of page 7 where the Examiner states "monitoring the hydrolysis (inactivation as claim) of a set of synthetic peptides…". The Examiner also states:

> The claim inactivation of a target protein involved with a disease or pathology in a mammal that ameliorate a disease is a property inherent or implicit to the method of Lien et al which uses the same compound in the method. [This is evidenct from the teachings of Shi et al at *e.g.*, paragraph [0003]. Shi states that members of the serine **protease** family which play important roles in a range of cellular functions and which have demonstrated causative roles in human **diseases** include tissue-type plasminogen activator and thrombin (thrombosis and blood clotting),

urokinase-type plasminogen activator (cancer and metastasis), trypsin and elastase (emphysema and liver **disease**) and angiotensin converting **enzyme** (hypertension).

Further, the Examiner states:

Furthermore, that the claim target protein is involved in different diseases or pathologies would be inherently accomplished by the methods of Lien since the same process steps comprising the same compounds is taguth by Lien (as evidenced by Shi).

Lien teaches at *e.g.*, page 73, col. 1:

> "...{i}t is useful to be able to generate proteases with new and desirable cleavage specificities. Such enzymes... could have practical applications in biotechnology... therapy (modulation of zymogen cascades, such as blood coagulation)...

Shi *et al.* teaches at *e.g.*, paragraph [003]:

> Members of the serine protease family which play important roles in a range of cellular functions and which have demonstrated causative roles in human diseases include tissue-type plasminogen activator and thrombin (thrombosis and blood clotting), urokinase-type plasminogen activator (cancer and metastasis), trypsin and elastase (emphysema and liver disease) and angiotensin converting enzyme (hypertension.

This rejection is respectfully traversed. The arguments directed to the rejection under 35 U.S.C. §102(b) as being anticipated by Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999)(as evidenced by Shi *et al.*, USP 20020197701) provided with the response mailed July 20, 2010 and the response mailed April 13, 2011 are incorporated by reference herein.

Preliminary Remarks

Applicant respectfully submits that the claims are amended herein to render it clear that the method results in the identification of a protease that cleaves a substrate sequence in a target protein such that the substrate sequence that is cleaved is different from a native substrate sequence of the wild-type scaffold protease. Hence, the method is a method whereby a protease is used as **a scaffold to engineer a protease with a new and different specificity from the wild-type** that exhibits cleavage activity for a substrate sequence in a target protein and that the substrate sequence is **different** from the substrate sequence cleaved by the unmutated scaffold protease. Further, the claim specifically requires that the cleavage of that sequence inactivate the target protein. Hence, because the target protein used in the method is one that is involved in a disease or pathology, the method permits identification of protease mutein that are candidate therapeutics because they can inactivate and destroy the functions in proteins that cause disease. As discussed below, Lien *et al.* does not disclose any method in which a protease is engineered to have cleavage activity for a new substrate

sequence nor is that substrate sequence one present in a target protein whose activity is to be inactivated.

Also, Applicant wishes to further clarify the Examiner's comment that hydrolysis is the same as inactivation as instantly claimed. As stated in previous responses of record, hydrolysis refers to the cleavage of peptide bonds by a protease. The result of hydrolytic cleavage by proteases can have diverse effects on target substrates, including inactivation of activities and activation of activities. Thus, hydrolysis does not always or necessarily result in the inactivation of an activity of a target protein. For example, C. Lopez-Otin and C.M. Overall summarize these activities as follows (*see e.g.* Nature Reviews:Molecular Cell Biology (2002) 3:809):

> Our view of the proteolytic universe has expanded considerably in recent years. Proteases were initially characterized as nonspecific degradative enzymes that are associated with protein catabolism. However, it is becoming increasingly recognized that proteolysis represents another mechanism for achieving precise cellular control of biological processes in all living organisms, through the highly specific hydrolysis of peptide bonds. This highly specific and limited substrate cleavage is termed proteolytic processing. Proteases, through their ability to catalyse irreversible hydrolytic reactions, regulate the fate and activity of many proteins by controlling appropriate intra- or extracellular localizations; shedding from cell surfaces; **activation or inactivation of proteases and other enzymes, cytokines, hormones or growth factors**; conversion of receptor agonists or antagonists; and exposure of cryptic neoproteins (which is when the proteolytic cleavage products are functional proteins with roles that are distinct from the parent molecule). [emphasis added by Applicant]

Further, Ehrmann and Clausen summarize at page 711 (Annu. Rev. Genet. (2004) 38:709-24):

> Not only is the activity of proteases subject to fine-tuned regulation but many proteases are themselves part of regulatory mechanisms. Many proteins including receptors, kinases, transcription factors, and structural components become modified by proteolysis in order to gain activity from a latent state or to alter an existing function. Furthermore, proteolysis is the most common mechanism for inactivating proteins. Thus, proteases play important roles by modulating enzyme activity and protein-protein interactions within multicomponent signaling pathways, thereby acting as master regulators.\

See also, Neurath *et al.* (1976) *Proc. Natl. Acad. Sci.*, 73:3825-3832.

Accordingly, proteolysis does not inherently lead to destruction and inactivation of target substrates, but rather, can also be involved in inducing activity of substrates. Hydrolysis of peptide bonds results in cleavage of a target; it does not describe whether the cleaved protein is inactive or active. In fact, cleavage of products can include activation cleavage or inactivation cleavage. In biological processes many proteases **activate** an

activity of a target protein. This occurs, for example, in proteolytic cascades, such as is involved in complement activation, coagulation and apoptosis, whereby the protease acts as a catalyst to **activate** downstream targets. For example, thrombin cleaves by hydrolysis the target protein substrates involved in the coagulation pathway . In particular, the downstream activated targets of thrombin are described in Bode *et al.* (1992) Protein Science, 1:426:471 as follows:

> α-Thrombin converts fibrinogen into fibrin, which consequently aggregates and forms the interconnecting network of thrombi. Furthermore, thrombin is able to **activate** several coagulation and plasma factors, such as Factor V, Factor VII, XIII, and protein C. [emphasis added]

Thus, many proteases that are involved in biological pathways act to activate an activity of a target protein, which is the opposite from the function of inactivating an activity of a target protein as instantly claimed. For example, in the coagulation cascade or complement pathway, proteases cleave target proteins in order to activate downstream effector molecules. This is completely different from the instant method, whose purpose is to modify protease molecules to develop therapeutic agents that destroy and inhibit the function of proteins adversely associated with disease. The end point of the instant method is not simply a protease that has improved activity and/or specificity, but rather is the identification of a modified protease that has a new specificity such that it destroys the activity of a target protein involved in a disease or pathology such that the modified protease can be used as an intervention of the activity of the target protein in order to serve as a candidate treatment for the disease or pathology. As discussed below, this is an affirmative feature of the instant method (*see e.g.* step (e), which requires testing the protease for inactivation of an activity of the target protein), such that the resulting identified proteases are those that inactivate an activity of a target protein, and in particular a target proteins whose activity results in a or causes a disease or pathology. It respectfully is submitted that the affirmative step (e) of the method of testing for inactivation of an activity of the identified protease distinguishes the method over general protease screening methods, since proteases are identified that exhibit a particular feature that is not common to all proteases. Also, it permits the identification of proteases that can destroy the function of a target protein that is associated with the pathogenesis of a disease or disorder, and thus the identified protease is a candidate therapeutic to treat the disease or disorder. Even if proteases inherently hydrolyze proteins when contacted with them at high enough concentrations and sufficient times under particular conditions, the cited art does not teach changing the specificity of the protease to cleave a particular sequence in a target that results in inactivation. Such hydrolysis to which

Examiner refers is not a new engineered and specific cleavage activity, but is natively present. Such activity is not selected as one that would create a therapeutic protein that specifically cleaves a target to inactivate it for treatment of a disease or disorder.

Relevant Law

The case law related to anticipation is set forth in previous responses, incorporated by reference herein. The Examiner is reminded that novelty requires more than locating each element with the four corners of a single document, but also that the single prior art reference must disclose those elements "arranged as in the claim." *Net Moneyln v. Verisign,* 545 F.3d.1359, 1369 (Fed. Cir. 2008).

The Rejected Claims

The instant method is a method for identifying from among a library of mutant proteases those that cleave and inactivate an activity of a target protein involved in a disease or pathology. The resulting identified proteases then can be candidate in vivo agents for treating the disease or pathology. For example, claim 1 recites:

> A method of producing and identifying a mammalian protease mutein that inactivates an activity of a target protein, wherein:
>
> the target protein is a protein that is involved with a disease or pathology in a mammal, whereby inactivating an activity of the target protein can treat the disease or pathology; and
>
> the method comprises the steps of:
>
> (a) producing a library comprising protease muteins comprising mutation(s) of a serine protease scaffold and/or biologically active portions thereof at an amino acid position selected from among amino acid residues 97, 98, 99, 100, 171, 174, 180, 189, 190, 191, 192, 215, 217, 218 , 226 and amino acid residues between amino acid 58 and amino acid 64, inclusive, whereby numbering of amino acid residues is by chymotrypsin numbering;
>
> (b) contacting members of the library with a target protein or with a polypeptide comprising a substrate sequence that is present in the target protein, wherein:
>
> the target protein is a cell surface molecule that transmits an extracellular signal for cell proliferation, a cytokine, a cytokine receptor or a signaling protein that regulates apoptosis; and
>
> cleavage of a substrate sequence in the target protein inactivates an activity of the target protein; and
>
> (c) measuring a cleavage activity and/or substrate specificity of at least two members of the library for the target protein or the polypeptide comprising the substrate sequence that is present in the target protein;
>
> (d) identifying protease mutein(s) from the library that have an increased cleavage activity and/or altered substrate specificity for cleaving the substrate sequence or the target protein, relative to the wild-type mammalian protease scaffold based on the measured activity and/or specificity in step (c); and

(e) testing the identified protease mutein(s) or biologically active portion thereof for cleavage and inactivation of an activity of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:

the protease mutein or biologically active portion thereof cleaves a substrate sequence in the target protein that is different from a native substrate sequence of the wild-type mammalian protease scaffold; and

the protease mutein inactivates an activity of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

Independent claims 53, 59 and 63 recite similar steps, but recite different further steps or different aspects of the scaffold protease or target protein. Dependent claims recite particulars of the method including the number of mutations contained in the protease muteins, the particular protease scaffold or target protein and other steps of the method.

For example, independent claim 53 recites that the mammalian protease scaffold is selected from granzyme A, granzyme B, granzyme M, cathepsin, trypsin, chymotrypsin, subtilisin, MTSP-1, elastase, chymase, tryptase, collagenase, papain, neutrophil elastase, complement factor serine proteases, ADAMTS13, neural endopeptidase/neprilysin, furin, cruzain, or urokinase plasminogen activator (uPA). Independent claim 53 also recites further steps of the method set forth as (f) to (i).

Independent claim 59 recites that the scaffold protease is a human protease and is granzyme A, granzyme B, granzyme M, cathepsin, trypsin, chymotrypsin, subtilisin, MTSP-1, elastase, chymase, tryptase, collagenase, papain, neutrophil elastase, complement factor serine proteases, ADAMTS13, neural endopeptidase/neprilysin, furin, cruzain, or urokinase plasminogen activator (uPA).

Independent claim 63 recites that the scaffold protease is a human protease and is granzyme A, granzyme B, granzyme M, cathepsin, MTSP-1, elastase, chymase, tryptase, chymotrypsin, collagenase, factor Xa, Protein C, plasma kallikrein, plasmin, trypsin, thrombin, complement factor serine proteases, papain, ADAMTS13, endopeptidase, furin, cruzain or urokinase plasminogen activator (uPA). Independent claim 63 also recites that the target protein is a caspase 3, tumor necrosis factor, tumor necrosis factor receptor, interleukin-1, interleukin-1 receptor, interleukin-2, interleukin-2 receptor, interleukin-4, interleukin-4 receptor, interleukin-5, interleukin-5 receptor, interleukin-12, interleukin-12 receptor, interleukin-13, interleukin-13 receptor, p-selectin, p-selectin glycoprotein ligand, Substance P, Bradykinin, PSGL, factor IX, immunoglobulin E, immunoglobulin E receptor, CCR5, CXCR4, glycoprotein 120, glycoprotein 41, hemagglutinin, respiratory syncytium virus fusion protein, B7, CD28, CD2, CD3, CD4, CD40, vascular endothelial growth factor,

VEGF receptor, fibroblast growth factor, endothelial growth factor, EGF receptor, TGF receptor, transforming growth factor, Her2, CCR1, CXCR3, CCR2, Src, Akt, Bcl-2, BCR-Abl, glucagon synthase kinase-3, cyclin dependent kinase-2 (cdk-2), or_cyclin dependent kinase-4 (cdk-4).

All claims, however, are directed to a method that require certain steps as follows:

1. **producing a library of mutant proteases** where members of the library contain a mutation or mutation compared to a scaffold protease at any one or more of the recited positions
2. **contacting** members of the library with a target protein or substrate sequence present in the target protein
3. **measuring the activity and/or specificity** of members of the library for the target protein or substrate sequence
4. **identifying** members that have **increased** activity and//or specificity
5. **testing the identified protease for cleavage and inactivation of an activity** of the target protein to verify that is function has been inactivated by the cleavage event, wherein the protease mutein **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the scaffold and cleavage of the the protease mutein ***at the new substrate cleavage sequence*** **inactivates an activity** of the target protein.

In particular, the instant claim recites an ***affirmative step*** of assaying or testing an identified protease mutant for inactivation of an activity of a target protein to verify the activity of the target protein is inactivated or reduced upon cleavage by protease, thereby selecting only mutant proteases that inactivate an activity of the target protein by cleavage of a new cleavage site for which the protease is mutated to cleave.

Analysis

The Examiner is reminded that anticipation requires the disclosure in a single prior art reference of each element of the claim under consideration. Also, **anticipation requires not only that each element must be disclosed in a single prior art reference, but also that the reference must disclose the elements of the claim as set forth in the claim**. On cannot pick and choose elements and combine them as claimed; they must be so-disclosed. Specifically, in the rejection as set forth, the Examiner provides basis and support in Lien *et al.* for various features of dependent claims (*see e.g.*, bottom of page 8 to page 9 of the Office Action). This support, however, is set forth at disparate sections in the disclosure of Lien *et al.* Lien *et al.* is an amalgamation of various methods from the literature, since it is a review article. Thus, the Examiner appears to be piecing apart aspects of various methods and combining them in order to achieve the method as instantly claimed. In addition, as discussed below, even if one could rely on Lien *et al.* in such a manner, all of the claimed elements are not disclosed in Lien *et al.*, even as parts of different methods.

There is no disclosure in Lien *et al.*, alone or as evidenced by Shi *et al.*, of any method in which a protease is engineered to have cleavage activity for a new and different substrate sequence that is different from a native substrate sequence of the scaffold protease nor where cleavage of such new substrate sequence results in inactivation of a target protein. Lien *et al.* also does not disclose any method, and in particular any method arranged as instantly claimed, that includes a step of "testing the identified protease for cleavage and inactivation of an activity of the target protein" to identify mutant proteases that inactivate an activity of the target protein. This step of "testing for inactivation of an activity" is an affirmative step, and hence is not an inherent property of the mutant proteases or any protease. If inactivation of an activity of a target protein was an inherent property of proteases, an affirmative step in the method a step of assaying to confirm that the function of the target protein has been inactivated would be pointless. It is not. It is not a step disclosed in Lien *et al.* nor can it be "inherent" since it requires contacting the protease with a target to confirm that the target is inactivated. In addition, as amended for clarity, the claim tequires that "the protease mutein or biologically active portion thereof cleaves a substrate sequence in the target protein that is different from a native substrate sequence of the wild-type mammalian protease scaffold.

As discussed above, the Examiner appears to be equating cleavage and hydrolysis with inactivation of an activity of a protein by virtue of specific cleavage of a new substrate sequence in a target. It respectfully is submitted that this is not correct.

Shi *et al.* (U.S. Patent application No. 20020197701)

Shi *et al.* is not relevant to the instant claims because it is not directed to any method of identifying protease muteins, in particular protease muteins that inactivate an activity of a target protein involved in a disease or pathology, nor any method including the recited steps as claimed. Shi *et al.* discloses the isolation of a native airway Trypsin-like protease molecule designated BlyP-2. It appears that the Examiner continues to cite Shi *et al.* as evidence that scaffold proteases are naturally involved in disease processes. It respectfully is submitted, however, that this is not relevant to the instant claims, which are directed to specific steps and methods of identifying a protease mutein that exhibits increased activity and/or specificity for a target protein compared to a scaffold protease, and that cleaves and inactivates an activity of the target protein because the protease is a mutein that is engineered to have a new cleavage sequence specificity, where such sequence is present in the target and where cleavage of such specific sequence inactivates the target. The role of natural proteases in diseases processes is not relevant because, as discussed above, the mechanism in which

proteases are involved in effecting such functions are diverse and can include inactivation or activation of target substrates. Natural proteases have not been engineered to specifically cleave as substrate sequence present in a target, where the substrate sequence is not a native cleavage site.

Further, the instant claims are directed to identifying **protease muteins that cleave and inactivate target proteins, where the target protein, not the scaffold protease is involved in a disease or disorders**. The instant claims are not directed to a method of using any protease to treat any disease or disorder. The instant claims are directed to methods of identifying mutein proteases that have new specificities such that they cleave a sequence in a target protein.

Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999)

Lien *et al.* is a review article that describes various strategies in the art to identify or produce mutant proteases, including rational mutagenesis and random mutagenesis (*see e.g.* introductory paragraph at page 73). In particular, Lien *et al.* discloses a targeted random mutagenesis strategy by randomizing several sites of a protease presumed to be involved in substrate binding to identify proteases with altered cleavage capabilities. For example, Lien *et al.* discloses methods of targeted combinatorial mutagenesis (TCM) by mutagenesis of residues that are specificity determinants (see *e.g.* at page 75, beginning at right column under heading "Specificity Determinants in Protease Substrate-Binding Sites"). Lien *et al.* also discloses strategies for generating libraries of mutant proteases, and the use of those protease libraries in screening and selection methods to identify new proteases (see *e.g.* at page 76, beginning at right column under heading "Altered Substrate Preferences in TCM libraries: Selection versus Screening.").

Lien *et al.* does not disclose any method that includes all elements as arranged in the instant claims. For example, Lien *et al.* does not disclose any method in which 1) cleavage activity and/or substrate sequence to a target protein is measured and a protease identified that exhibits **increased** cleavage activity and/or specificity for a new cleavage sequence present in a target protein; 2) the protease mutein is **tested for cleavage and inactivation of an activity** of the target protein; and 3) identification of a protease mutein that cleaves **a substrate sequence that is different** from a native substrate sequence of the protease scaffold, whereby cleavage of the substrate sequence in the target protein inactivates an activity of the target protein.

While Lien *et al.* states that mutant proteases that are identified to have new and desirable cleavage specificities can be used in therapy, for example for the modulation of

zymogen cascades, such as blood coagulation, Lien *et al.* does not disclose ***any method*** that includes all steps as instantly claimed, nor any method to achieve the results of a protease mutein that exhibits the properties as instantly claimed (*i.e.* the ability to cleave a substrate sequence that is different than a native substrate sequence, whereby cleavage of the substrate sequence inactivates an activity of the target protein). There is no disclosure in Lien *et al.* of any method of identifying a protease that inactivates an activity of a target protein such that the function of the target protein is destroyed.

The methods disclosed in Lien *et al.* are directed to various methods of generating combinatorial mutant libraries (see *e.g.* at pages 74 to middle of second column on page 76) and various methods of screening or selection of the protease library in order to identify active protease mutants (see *e.g.* beginning middle of second column on page 76 to page 77). With respect to disclosure of the target substrate used in any method disclosed in Lien *et al.*, Lien *et al.* does not specifically disclose any in which proteases are screened for cleavage activity for a substrate sequence that is different from a native substrate sequence, where cleavage of the substrate sequence inactivates an activity of the target protein. Lien *et al.* is primarily directed to methods of improving the specificity/selectivity for **existing** substrate sequences of proteases.

In addition, and as discussed above, the instant method is different from any method set forth in Lien *et al.* because the instant method includes an **affirmative step** of "***testing the identified protease(s) for cleavage and inactivation of an activity of the target protein that contains the substrate sequence.***" Hence, the instant method includes a step of testing the activity of the target protein in the presence of a protease mutant identified in step d) as having increased cleavage activity for the substrate sequence, and in this secondary screen only identifying those protease mutants that inactivate (*e.g.* destroy or reduce) the activity of the target protein. Lien *et al.* does not include a method that includes a step of testing the activity of the target protein in the presence of a protease mutein to verify that its function has been destroyed.

Notwithstanding that the recited steps as arranged in the method of independent claims 1, 53, 59 and 63 are not disclosed in Lien *et al.*, Lien *et al.* also does not disclose other particular aspects and features of the independent claims or dependent claims. For example, a review of Lien *et al.* shows that it does not disclose or even mention any **target protein** that is cleaved by a protease in the method, where the target is caspase 3, tumor necrosis factor, tumor necrosis factor receptor, interleukin-1, interleukin-1 receptor, interleukin-2, interleukin-2 receptor, interleukin-4, interleukin-4 receptor, interleukin-5, interleukin-5

receptor, interleukin-12, interleukin-12 receptor, interleukin-13, interleukin-13 receptor, p-selectin, p-selectin glycoprotein ligand, Substance P, Bradykinin, PSGL, factor IX, immunoglobulin E, immunoglobulin E receptor, CCR5, CXCR4, glycoprotein 120, glycoprotein 41, hemagglutinin, respiratory syncytium virus fusion protein, B7, CD28, CD2, CD3, CD4, CD40, vascular endothelial growth factor, VEGF receptor, fibroblast growth factor, endothelial growth factor, EGF receptor, TGF receptor, transforming growth factor, Her2, CCR1, CXCR3, CCR2, Src, Akt, Bcl-2, BCR-Abl, glucagon synthase kinase-3, cyclin dependent kinase-2 (cdk-2) and cyclin dependent kinase-4 (cdk-4) (see *e.g.*, claims 12, claim 62 and claim 63)).

Therefore, Lien *et al.* does not disclose all elements of the methods **as claimed**. Thus, Lien *et al.* does not anticipate any pending claim. Accordingly, for at least these reasons, Lien *et al.* does not anticipate independent claims 1, 53, 59 and 63, or any of the dependent claims therefrom.

V. THE REJECTION OF CLAIMS 1, 3-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63 and 65-78 UNDER 35 U.S.C. §103

Claims 1, 3-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63 and 65-78 are rejected under 35 U.S.C. §103(a) as being unpatenable over Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999, 2:73-90) (as evidenced by Shi *et al.*, USP 20020197701). The Examiner states that Lien *et al.* teaches a method using combinatorial mutagenesis that includes producing libraries of proteases containing mutations, contacting the library with a substrate and screening and selecting a protease.

Applicant respectfully traverses this rejection. The Preliminary Remarks above under 35 U.S.C. §102(b) as being anticipated by Lien *et al.* are incorporated herein by reference.

Relevant Law

The case law related to obviousness is set forth in previous responses. In particular, unexpected properties must always be considered in the determination of obviousness. A compound's structure and properties are inseparable so that unexpected properties are part of the subject matter as a whole. In re Papesch, 315 F.2d 381, 391, 137 USPQ 43, 51 (CCPA 1963)

The disclosure of the applicant cannot be used to hunt through the prior art for the claimed elements and then combine them as claimed. In re Laskowski, 871 F.2d 115, 117, 10 USPQ2d 1397, 1398 (Fed. Cir. 1989). "To imbue one of ordinary skill in the art with knowledge of the invention in suit, when no prior art reference or references of record convey or suggest that knowledge, is to fall victim to the insidious effect of a hindsight syndrome

wherein that which only the inventor taught is used against its teacher" W.L. Gore & Associates, Inc. v. Garlock Inc., 721 F.2d 1540, 1553, 220 USPQ 303, 312-13 (Fed. Cir. 1983).

In addition, the concept of inherency is not applicable to the question of obviousness. In re Sporman, 363 F.2d 444, 150 USPQ 449 (CCPA 1965). To state that a reference inherently teaches an element of a claim rebuts any finding of prima facie obviousness. The concept of inherency is not properly applicable to the question of obviousness (see, In re Sporman, 363 F.2d 444, 150 USPQ 449 (CCPA 1965)).

Obviousness and inherency are entirely different questions; that which may be inherent is not necessarily known and, therefore, is an indication of unobviousness (In re Sporman, 363 F.2d 444, 449, 150 USPQ 449, 452 (CCPA 1965; see, also In re Naylor, 360 F.2d 765, 152 USPQ 106 (CCPA 1966); In re Adams, 356 F.2d 998, 148 USPQ 742 (CCPA 1966); and In re Shetty, 566 F.2d 81, 195 USPQ 753 (CCPA 1977)).

Reference to an unexpected property as inherent begs the question of whether an unexpected property rebuts prima facie obviousness. The unexpected property is part of the invention as a whole, and, therefore, evidence of unobviousness of the claimed subject matter. In re Naylor a process for preparing a polybutadiene polymer having unexpected properties was at issue. The CCPA held that the fact that a rubbery polybutadiene having high 1, 2-addition might be inherent in following the combined teachings of the prior art is immaterial, if one of ordinary skill in the art would not appreciate or recognize that inherent result. In re Adams, the CCPA held that since properties of a claimed structure are always inherent, it is "transparently erroneous" to state that subject matter cannot be patented on the basis of an inherent property. In re Shetty, the court held that "inherency is quite immaterial if, as record established here, one of ordinary skill in the art would not appreciate or recognize that inherent result."

The Rejected Claims

The rejected claims are set forth above

Analysis

Preliminary Remarks

It respectfully is submitted that the instant method is based on the recognition that proteases can be used as scaffolds to engineer candidate therapeutics that have specificity for particular target proteins known to cause disease, whereby inactivation of the target protein interrupts a pathway involved in a disease or pathology. The protease scaffold used in the method is a reagent that is modified to recognize a new and different substrate sequence from

a native substrate sequence recognized by the scaffold , where the sequence is present in a desired target protein involved in a disease or pathology, whereby cleavage of the substrate sequence inactivates an activity of the target protein. For example, the instant specification states:

> The invention is broadly drawn to the modification of proteases to alter their substrate sequence specificity, so that the modified proteases cleave a target protein which is involved with or causes a pathology. [emphasis added].

And, also states:

> In some examples, the engineered protease is designed to cleave any of the target proteins in Table 1, thereby inactivating the activity of the protein. The protease can be used to treat a pathology associated with that protein, by inactivating one of the target proteins.

It respectfully is submitted that none of the methods taught or suggested in Lien *et al.*, singly or in any combination with Shi *et al.*. Neither reference, singly or in any combination teaches or suggest any method that results in the identification of a protease mutein that inactivates an activity and destroys a function of a target protein, whereby the target protein is one in which inactivation of an activity thereof can treat the disease or pathology. None of the methods achieves the results of the instant method of a protease mutein that exhibits the properties as instantly claimed (*i.e.* the ability to cleave a substrate sequence that is different from a native substrate sequence, whereby cleavage of the substrate sequence inactivates an activity of the target protein).

Differences Between the Claims and Teachings of the Cited References

Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999)

Regarding protease screening methods, Lien *et al.* teaches that proteases can be engineered for academic, industrial and medical applications. Lien *et al.* reviews various applications based on quantitative assessments and distinguishes screening methods from selection methods (see *e.g.* pp. 76-82). The focus of Lien *et al.* is on the use of targeted combinatorial mutagenesis (TCM) in the screening methods. In particular, Lien *et al.* reviews the results of various methods in terms of elucidating the roles of various residues in altering the cleavage activity or substrate specificity of the protease. Lien *et al.* provides examples where the strategy has been employed to identify inactive enzymes, to identify proteases with increased activity for an existing substrate, and to identify proteases with altered substrate specificity. Lien *et al.* summarizes the results of many of these studies in terms of elucidating the plasticity, tolerance and/or amenability of individual residues in proteases involved in altering the activity and/or specificity of particular proteases.

Hence, Lien *et al.* is merely a review article summarizing that proteases can be evolved to have altered activities and specificities for various substrates and for various applications. As discussed above, Lien *et al.* does not teach or suggest any method including all limitations as instantly claimed. For example, Lien *et al.* **does not** teach or suggest a method that includes a step as claimed of:

> (d) identifying protease mutein(s) from the library that have an **increased** cleavage activity and/or altered substrate specificity for cleaving the substrate sequence or the target protein, relative to the wild-type mammalian protease scaffold based on the measured activity and/or specificity in step (c); and
>
> (e) **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

Among the methods reviewed by Lien *et al.* using targeted combinatorial mutagenesis (TCM), none teaches or suggests a method of identifying a mutein protease that has increased specificity and/or cleavage activity for a substrate sequence in a target protein, whereby the protease mutein cleaves a substrate sequence that is different from a native substrate sequence of the scaffold protease, and cleavage of the substrate sequence in the target protein inactivates an activity of the target protein. There is no teaching or suggestion of any method that includes a step of testing the identified protease for cleavage and inactivation of an activity of the target protein. There is no teaching or suggestion providing any reason to identify protease muteins that inactivate a target protein involved in a disease or pathology, or any method of doing so, or that doing so could identify therapeutic agents.

The Examiner states that the methods in Lien *et al.* would inherently produced the desired result. For example, the Examiner states that Lien *et al.* teaches (*e.g.* page 73, col 1):

> ...{i}t is useful to be able to generate proteases with new and desirable cleavage specificities. Such enzymes... could have practical applications in biotechnology... therapy (modulation of zymogen cascades, such as blood coagulation)...

First, there is an affirmative method requiring testing of the protease to confirm that it cleaves the target protein and inactivates it. Second, the concept of inherency has no place in a consideration of obviousness; what did not previously exist cannot be said to be inherent. Applicant respectfully submits that inherency is not relevant to a finding of nonobviousness in this case, where the property achieved by the instant method (*i.e.* the ability to cleave a substrate sequence that is different than a native substrate sequence, whereby cleavage of the substrate sequence inactivates an activity of the target protein) is not taught or suggested in the cited reference. Lien *et al.* does not provide any teaching that a protease mutein that specifically inactivates a target protein that is involved in a disease or pathology would have any therapeutic applications. Indeed, as discussed above, proteases have diverse functions that include activation and inactivation of target proteins. In particular, zymogen cascades, as taught by Lien *et al.*, are specifically associated with proteolytic activation cleavage events.

The only teaching in Lien *et al.* with respect to proteases involved in blood coagulation or any therapeutic application is a summary of the results of a study that teaches use of a mutagenesis method to identify a mutant of thrombin with an altered substrate preference from fibrinogen to protein C, thereby converting the mutant thrombin (*activating it)* into a potent anticoagulant (see at page 73, last sentence, first paragraph; summarizing Tsiang *et al.* (Biochemistry (1996) 3:16449)). It is known to one of ordinary skill in the art that protein C and fibrinogen are zymogens, and that a zymogen is a protease that exhibits activity following cleavage by a protease. Lien *et al.* is describing *activation cleavage* in which cleavage by a protease converts a zymogen into an active form. Thus, for example, it is known to one of skill in the art that both fibrinogen and protein C are natural substrates of thrombin in the proteolytic cascade, whereby proteolytic events, by a process known as **activation cleavage, activate** downstream targets (see *e.g.* Tsiang *et al.*). For example, thrombin activates and converts fibrinogen to fibrin and thereby catalyzes coagulation-related reactions. Thrombin also activates protein C, which is an inhibitor of the coagulation cascade. Therefore, a thrombin mutant that has greater substrate preference for protein C will exhibit greater anticoagulant activity.

Thus, in the example in Lien *et al.* summarizing the engineering of a protease for therapeutic applications in blood coagulation, Lien *et al.* teaches (through the teachings of Tsiang *et al.*) a method to select a protease mutein that exhibits greater activation cleavage activity for its native substrate protein C compared to its other native substrate fibrinogen in order to **activate** the protein. The thrombin mutants are identified based on their activation of protein C, and not on inactivation of protein C. This is completely different from the instant

method, and in fact is the **opposite effect** of the results achieved by the instant method in which proteases are identified that cleave a substrate sequence that is **different** than a native substrate sequence, whereby cleavage of the substrate sequence **inactivates** an activity of the target protein.

Further, in the example in Lien *et al.* directed to identification of the thrombin mutant (citing to Tsiang *et al.*), the mutant was identified because it exhibited altered selectivity for protein C compared to fibrinogen, and **not** because it exhibited **increased** cleavage activity for protein C compared to the scaffold protease not containing the mutation. For example, the study by Tsiang *et al.* as cited in Lien *et al.* selected mutants based on the ratio of the specific activity of PC activation (PA) over the specific activity of fibrinogen clotting (FC). The ratio for wildtype was valued at 1, and mutants were identified with a ratio of greater than 20. Thus, the method for identifying a protease for therapy applications as taught in Lien *et al.* was based in identifying one that exhibited changes in selectivity for native target proteins, and not increased cleavage and/or specificity for a substrate sequence in a target protein that is different from the native substrate sequence.

Therefore, with respect to a method directed to identification of proteases for use in therapy, the only teaching in Lien *et al.* is directed to a method that differs in several aspects compared to the instant method. Lien *et al.* does not teach or suggest a method of identifying a proteases mutein that cleaves and inactivates a target protein involved in a disease or pathology, whereby the mutein exhibits increased cleavage activity for the protein compared to the scaffold and cleaves a substrate sequence that is different from the scaffold, whereby cleavage of the substrate sequence inactivates an activity of the target protein.

Shi *et al.* (U.S. Patent application No. 20020197701)

Shi *et al.* does not cure the deficiencies in the teachings of Lien *et al.* Shi *et al.* is directed to the isolation of a native airway Trypsin-like protease molecule designated BlyP-2. Shi *et al.* does not teach or suggest any methods of engineering scaffold protease, nor any method of identifying a protease mutein that inactivates an activity of the target protein as claimed. Further, as discussed above, the teachings in Shi *et al.* that proteases are naturally involved in disease proteases is completely irrelevant to the instant method. The instant method is a method of identifying a protease mutein to inactivate a target protein involved in a disease or pathology. Hence, the protease mutein is a candidate therapeutic for treating a disease or pathology, and is not involved in any natural disease processes.

Conclusion

The Examiner has failed to set forth a prima facie case of obviousness because neither Lien *et al.* nor Shi *et al.*, singly or in combination, teaches all elements as claimed. As discussed above, Lien *et al.* fails to teach or suggest all elements of the claimed method, and particular features of dependent claims. Importantly, Lien *et al.* fails to teach or suggest any method that includes the step of:

> (d) identifying protease mutein(s) from the library that have an **increased** cleavage activity and/or altered substrate specificity for cleaving the substrate sequence or the target protein, relative to the wild-type mammalian protease scaffold based on the measured activity and/or specificity in step (c); and
>
> (e) **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

Shi *et al.* is not directed to methods of identifying a protease mutein or screening of protease muteins, but is directed to the discover of a new natural serine protease. Hence, Shi *et al.* does not cure this deficiency in the teachings of Lien *et al.*, since Shi *et al.* also fails to teach or suggest the above limitations, and other limitations and features in the claims and dependent claims. Accordingly, since *prima facie* obviousness requires that the references, singly or in any combination, teach or suggest all elements as claimed, the Examiner has failed to set forth a case of *prima facie* obviousness.

VI. THE REJECTION OF CLAIMS 1-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63, AND 65-78 UNDER 35 U.S.C. §103

Claims 1-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63, and 65-78 are rejected under 35 U.S.C. §103(a) as being unpatenable over Lien *et al.* (Combinatorial Chemistry and High Throughput Screeing, 1999, 2:73-90) in view of either Harris *et al.* I (J. Biol. Chem., 273:27364-27373 (1998) or Harris *et al.* II (Current Opinion in Chemical Biology, 2:127-132 (1998)) and Waugh *et al.* (Nature Structure Biology). The Examiner states that while Lien *et al.* does not disclose the enzyme as granzyme and the substrate as caspase, Harris *et al.* I teaches the substrate specificity of granzyme B and that caspases are likely substrates, and teaches that Arginine 192 is a structural determinant of specificity of granzyme B; Harris *et al.* II teaches the same method as Harris *et al.* I; and Waugh *et al.* teaches that granzymes are a vital component of the ability to induce apoptosis. Thus, the Examiner concludes that it

would have been obvious to one of ordinary skill to use the serine protease granzyme and the target protein caspase in the method of Lien *et al.* The Examiner concludes that:

> The claimed methods are routine steps in screening and identifying a specific mutant from a library of mutants that binds to a specific target protein to identify a protease mutant with improved (*e.g.* increased) property against its target. The claim method is nothing more than a predictable result expected of the method of screening protease muteins against a target protein.

Applicant respectfully traverses this rejection. This rejection is similar to previous rejections made of record that have been addressed in responses filed August 7, 2008, June 11, 2009, July 20, 2010 and April 13, 2011, each of which are incorporated by reference herein. Further, the Preliminary Remarks set forth above in the rejection under 35 U.S.C. §102(b) as being anticipated by Lien *et al.* and 35 U.S.C. §103(a) as being unpatenable over Lien *et al.* in view of Shi *et al.* also is incorporated by reference herein.

Relevant Law

The case law related to obviousness is set forth in previous responses of record, each incorporated by reference herein. Case law is also set forth above.

The Rejected Claims

The rejected claims are set forth above.

Analysis

Differences Between the Claims and Teachings of the Cited References

Lien *et al.* (Combinatorial Chemistry and High Throughput Screening, 1999)

As discussed above, Lien *et al.* does not teach or suggest a method including all limitations as instantly claimed. For example, Lien *et al.* **does not** teach or suggest a method that includes steps as claimed of:

> (d) identifying protease mutein(s) from the library that have an **increased** cleavage activity and/or altered substrate specificity for cleaving the substrate sequence or the target protein, relative to the wild-type mammalian protease scaffold based on the measured activity and/or specificity in step (c); and
>
> (e) **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

None of Harris *et al.* I, Harris *et al.* II or Waugh *et al.*, singly or in any combination, cures these defects.

Harris *et al.* I (J. Biol. Chem., 273:27364-27373 (1998)

Harris *et al.* I is directed to a study of the mechanism by which granzyme B mediates apoptosis of target cells and demonstrates that granzyme B displays extended substrate specificity and proposes a model by which granzyme B acts. In this context, for example, while Harris *et al.* I teaches two mutants of granzyme B at amino acid residue 192, Harris *et al.* teaches that mutation of amino acid residue 192 in granzyme B **reduces** its hydrolysis of an optimal tetrapeptide substrate. Thus, Harris *et al.* does not teach or suggest employing granzyme B nor any protease as a scaffold in a method to identify any protease mutants that exhibit increased cleavage activity and/or substrate specificity for a target protein involved in a disease or pathology as required by the instant claims.

Harris *et al.* I also does not teach or suggest any method that includes as a step testing the identified protease mutant for inactivation of an activity of the target protein. There is no teaching or suggestion in Harris *et al.* of a step of:

> **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

In this regard, Harris *et al.* teaches that caspase is a target substrate of granzyme B. Like the coagulation pathway, the apoptosis pathway involves a proteolytic cascade, whereby proteases act as catalysts to **activate** downstream molecules. For example, granzymes such as granzyme B cleave caspases, thereby altering the conformation and leading to their *activation*. Specifically, granzyme B cleaves procaspase-3 to its activated caspase-3 form, which itself performs further downstream functions in the apoptosis pathway. For example, Harris *et al.* I teaches at page 27364, 2nd column:

> Although granzyme B is the only known mammalian serine protease to have P1-proteolytic specificity, it is shared with the caspases, a family of cysteine proteases that are also activated during apoptosis. The link between granzyme B and the caspases has been strengthened by studies indicating that **granzyme B can cleave and activate certain members of the caspases**, and it has been

> suggested that this is one of the mechanisms by which granzyme B mediates apoptosis in vivo. [emphasis added]

Harris *et al.*, thus, teaches that cleavage by granzyme activates its target protein.

Harris *et al.* II (Current Opinion in Chemical Biology, 2:127-132 (1998))

Harris *et al.* II is a review article describing determinants of protease specificity and methods for modifying specificity as a means to "understand the factors involved in protease specificity." Harris *et al.* II teaches that several methods that have been employed to make changes in enzyme specificity in order to understand protein design principles. The first method, designated rational redesign specificity, requires a detailed understanding of the catalytic mechanism and sequence determinants for a particular protease as basis for predictably altering specificity. Harris *et al.* II also teaches complete random mutagenesis and states that this requires large libraries of mutants to identify a desired function. Another method, employs comparative analysis using homologous proteins that differ in substrate specificity. In another method, Harris *et al.* describes the use of alanine scanning to identify residues involved in activity, followed by replacement of identified residues with all 19 amino acids to produce modified proteins with altered activity. In a final method, Harris *et al.* teaches using directed evolution, such as DNA shuffling methods, to generate enzymes with modified activities.

Harris *et al.* II provides a general review of the state of the art at the time of its publication. Harris *et al.* II is a review article that describes methods that have been used to identify structural determinants. None of the methods taught by Harris *et al.* II are for generating a protease that has specificity for a particular target involved in a pathology, and none include a step of identifying a protease that inactivates an activity of a target protein by:

> **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

Hence, Harris *et al.* II fails to teach or suggest the elements of the instantly claimed methods, particularly those elements pertaining to the cleavage and inactivation of the target protein involved in a disease or pathology that are deficient in the other cited references.

Thus, the combination of Harris *et al.* II, singly and in combination with Harris *et al.* I and/or Lien *et al.*, fails to teach or suggest each of the elements of the instantly claimed methods.

Waugh *et al.* (Nature Structure Biology 7: 762-765 (2000)

Finally, Waugh *et al.* also is of little relevance to the instant claims. Waugh *et al.* teaches that granzymes are involved in inducing apoptosis by acting on downstream substrates such as caspases by **activation cleavage** (not inactivation as instantly required in the claimed method). Waugh *et al.* also teaches the elucidation of the molecular determinants of substrate specificity in granzyme B. In particular, the first paragraph of Waugh *et al.* teaches:

> Granzymes are a vital component of the cytotoxic lymphocyte's ability to induce apoptosis, contributing to rapid cell death of a tumor or virally infected target cell by the cleavage of downstream substrates and the **activating** cleavage of caspases. [emphasis added]

Thus, Waugh *et al.* describes that the normal function of granzyme is to activate caspases to induce apoptosis. Waugh *et al.* does not teach or suggest any protease that inactivates an activity of a target protein, in particular a target protein involved in a disease or pathology, nor a method of identifying such a protease mutein as instantly claimed. There is no teaching in Waugh *et al.* of a method that includes a step of:

> **testing** the identified protease mutein(s) or biologically active portion thereof for cleavage and **inactivation of an activity** of the target protein that contains the substrate sequence to verify that the function of the target protein has been inactivated by the cleavage event, wherein:
>
> the protease mutein or biologically active portion thereof **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold; and
>
> **cleavage of the substrate sequence inactivates an activity** of the target protein, thereby identifying a protease mutein or a biologically active portion thereof that inactivates an activity of a target protein that is involved with the disease or pathology.

There is no teaching or suggestion in Waugh *et al.* of any method of screening for mutant proteases to identify a protease that has increased cleavage activity and/or substrate specificity for a target protein involved in a disease or pathology that includes a step of testing the identified protease mutants for inactivation of an activity of the target protein, whereby the protease mutein cleaves a substrate sequence that is different from a native substrate sequence and cleavage of the substrate sequence inactivates an activity.

Further, while Waugh *et al.* teaches the elucidation of the molecular determinants of substrate specificity in granzyme B, Waugh *et al.* does not teach or suggest that mutation of any of these residues would alter substrate specificity, in particular substrate specificity for

a target protein involved in a disease or pathology. Waugh *et al.* does not teach or suggest any mutants of granzyme B. There is no teaching or suggestion in Waugh *et al.* of a method of generating mutants of granzyme B, such as a library of protease mutants as claimed that contain mutations at amino acid residues 99, 171, 174, 180, 189, 190, 191, 192, 215, 217, 218 ,226, or between 58 and 64, inclusive, by chymotrypsin numbering as recited the claims.

Thus, Waugh *et al.*, singly or in combination with Lien *et al.*, Harris *et al.* I or Harris *et al.* II, does not teach or suggest the claimed method.

Conclusion

Hence, the Examiner has failed to set forth a prima facie case of obviousness because none of Lien *et al.*, Harris *et al.* I, Harris *et al.* II and/or Waugh *et al.*, singly or in any combination thereof, teaches all elements as claimed. The independent claims are directed to a method of producing and identifying a protease mutein that cleaves and inactivates a target protein involved in a disease or pathology, including steps of:

1. **producing a library of mutant proteases** where members of the library contain a mutation or mutation compared to a scaffold protease as a recited amino acid residue
2. **contacting** members of the library with a target protein or substrate sequence present in the target protein
3. **measuring the activity and/or specificity** of members of the library for the target protein or substrate sequence
4. **identifying** members that have **increased** activity and//or specificity; and
5. **testing** the identified protease mutein(s) for cleavage and **inactivation of an activity** of the target protein, whereby the protease mutein **cleaves a substrate sequence in the target protein that is different** from a native substrate sequence of the wild-type mammalian protease scaffold and **cleavage of the substrate sequence inactivates an activity** of the target protein,

In particular, the cited references, singly or in combination, do not teach or suggest a method that includes a step of 3) and 4) of measuring and identifying protease muteins that have increased cleavage activity and/or substrate specificity for a substrate sequence in a target protein or a target protein involved in a disease or pathology, or a step of 5) testing identified protease muteins for inactivation of an activity of the target protein. Since *prima facie* obviousness requires that the references, singly or in any combination, teach or suggest all elements as claimed, the Examiner has failed to set forth a case of prima facie obviousness.

Rebuttal to Examiner's Comments

1) The Examiner specifically rebuts Applicants summary of the results in the Examples that they are not the same as in Harris *et al.* It respectfully is submitted that this

summary was provided to show that the Examples demonstrate practice of the method as claimed, and identification of a protease mutein that cleaves a substrate sequence that is different from a native substrate sequence of the wildtype scaffold, whereby cleavage of the substrate sequence inactivates an activity of the target protein. It is this result, which Applicant stated was not taught or suggested in Harris *et al.* Indeed, as discussed above, Harris *et al.* teaches **activation** of caspases by granzyme B, and not their inactivation. It appears that the Examiner also appreciates that Harris *et al.* does not teach any method of identifying a protease that inactivates an activity of a target protein, since the Examiner states:

> Thus Harris teaches mutations of 192 to glutamate activate caspase 7 which is different from the caspase-3 disclosed in the specification that is inactivated by the single mutant granzyme B I99A/N218A. (The claims recite for several mutants). Since there are apparently several caspase, activation and inactivation appear to be sequence dependent, as taught in the reference above.

Further, with respect to the Examiner's comment above that inactivation and activation appear to be sequence dependent, it respectfully is submitted that the instant method specifically selects for those cleavage events that result in inactivation, such that the mechanism (*e.g.* "sequence dependence") of the inactivation is not relevant. For example, the end point of the instant method is the identification of a modified protease that destroys the activity of a target protein involved in a disease or pathology. This result of the method is achieved by an affirmative feature of the instant method (*see e.g.* step (e), which requires testing the protease for inactivation of an activity of the target protein), such that the resulting identified proteases are those that inactivate an activity of a target protein, and in particular a target proteins whose activity results in a or causes a disease or pathology. It respectfully is submitted that the affirmative step (e) of the method of testing for inactivation of an activity of the identified protease distinguishes the method over general protease screening methods, since proteases are identified that exhibit a particular feature that is not common to all proteases.

VII. THE REJECTION OF CLAIMS 1-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63, AND 65-78 FOR NONSTATUTORY OBVIOUSNESS-TYPE DOUBLE PATENTING

Claims 1-7, 9, 11-16, 45, 48, 51-53, 56-59, 61-63, and 65-78 are provisionally rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claims 1-10 of copending Application No. 12/005,949 ('949) application.

As stated in the responses of record, Applicant requests deferral of resolution of this issue. It is not possible to assess whether claims at allowance in each application will overlap requiring a terminal disclaimer until there is an indication of allowable subject matter in at least one application. It is premature to file a terminal disclaimer at this time. In this respect, The Examiner is reminded of MPEP 804(I)(B)(1), which recites, in part:

> If a 'provisional' nonstatutory obviousness-type double patenting (ODP) rejection is the only rejection remaining in the earlier filed of the two pending applications, while the later-filed application is rejectable on other grounds, the examiner should withdraw that rejection and permit the earlier-filed application to issue as a patent without a terminal disclaimer.

The instant application was filed on October 02, 2003, whereas, U.S. application Serial No. 12/005,949, is a later filed application filed on December 28, 2007. Since this application was filed four years before 12/005,949, it is highly unlikely that U.S. Application. Serial No. 12/005,949 could issue before the instant application. Thus, MPEP 804(I)(B)(1) applies. Further, if, when one or both applications is deemed allowed, it is determined that a terminal disclaimer is necessary, Applicant will file a terminal disclaimer.

* * *

Consideration of the above remarks, entry of this amendment and continued examination of the application on the merits respectfully are requested.

Respectfully submitted,

Stephanie Seidman
Reg. No. 33,779

Attorney Docket No. 33328.04905.US01/4905
Address all correspondence to: 13565
Stephanie Seidman
Mckenna Long & Aldridge LLP
4435 Eastgate Mall, Suite 400
San Diego, California 92122
Telephone: (619) 595-8010
Facsimile: (858) 595-8135
email: sseidman@mckennalong.com

Proc. Natl. Acad. Sci. USA
Vol. 73, No. 11, pp. 3825–3832, November 1976
Biochemistry

Role of proteolytic enzymes in biological regulation (A Review)*

(limited proteolysis/zymogen activation/control mechanisms)

HANS NEURATH AND KENNETH A. WALSH

Department of Biochemistry, University of Washington, Seattle, Wash. 98195

Contributed by Hans Neurath, September 7, 1976

ABSTRACT **Many enzymes, hormones, and other physiologically active proteins are synthesized as inactive precursors (zymogens) that are subsequently converted to the active form by the selective enzymatic cleavage (limited proteolysis) of peptide bonds. The ultimate agency of activating enzymatic function is limited proteolysis, either in a single activation step or in a consecutive series (cascade). The specificity of each activation reaction is determined by the complementarity of the zymogen substrate and the active site of the attacking protease. The sequence of consecutive activation reactions is regulated by the specificity of each enzyme, whereas the degree of amplification of the initial stimulus is determined by the efficiency of each activating step.**

Zymogen activation produces a prompt and irreversible response to a physiological stimulus, and is capable of initiating new physiological functions. Typical examples are the processes of blood coagulation, fibrinolysis, complement activation, hormone production, metamorphosis, fertilization, supramolecular assembly, and digestion. The zymogens of the pancreatic serine proteases, in particular, have served as models for detailed studies of the nature of the molecular changes that are involved in the dramatic increase in enzymatic activity that ensues upon limited proteolysis of the zymogen.

In recent years, it has become evident that many proteins are synthesized as inactive precursors or zymogens and that these are subsequently converted to physiologically active forms by the selective enzymatic cleavage of peptide bonds. This process is known as zymogen activation, a term which initially was applied to the activation of precursors of proteolytic enzymes such as trypsinogen, chymotrypsinogen, or procarboxypeptidase (1). It is now apparent that the same type of reaction is involved in a great variety of biological processes, such as blood coagulation, fibrinolysis, complement reaction, hormone production, development, differentiation, and supramolecular assembly, all of which involve zymogen activation in one or more steps (2–9). In the present article, we shall attempt to show that activation by limited proteolysis is indeed an important control element which can initiate new physiological functions or regulate preexistent ones.

Virtually all zymogen activation reactions require the enzyme-catalyzed cleavage of a unique peptide bond by "limited proteolysis." This term was first introduced by Linderstrøm-Lang and Ottesen (10) to describe the restrictive peptide bond cleavage that induces the conversion of ovalbumin to a different crystalline form, plakalbumin, under the influence of the bacterial protease subtilisin. Numerous examples of limited proteolysis have since been described and studied in detail, such as the tryptic conversion of chymotrypsinogen to chymotrypsin (1), the release by subtilisin of the amino-terminal segments of ribonuclease (11), and the conversion of proinsulin to insulin (5). Limited proteolysis is the last step in the synthesis of many biologically active proteins and probably the first step in protein degradation (12). The specificity of limited proteolysis is best understood in terms of the three-dimensional structure of a protein substrate and of the attacking protease because the region of the protein substrate containing the susceptible peptide bond must fit the active site of the attacking protease in order for amino acid residues of the substrate to interact with primary as well as secondary binding sites of the enzyme (13). In general, limited proteolysis is therefore directed toward surface loops and random segments of polypeptide chains rather than toward internal domains, helices, or pleated sheets.

The activation of zymogens usually occurs by proteolytic cleavage of a peptide bond in a region that is amino terminal relative to the active site of the protein. This may be a consequence of the process of protein biosynthesis, which proceeds in the direction from the amino to the carboxyl end. If it is assumed that the protein assumes its correct tertiary structure as regions of the polypeptide chain are synthesized, the zymogen will be formed prior to the enzyme. Were the activation peptide attached to the carboxyl end, trypsin would be synthesized before trypsinogen, fibrin before fibrinogen, or collagen before procollagen. By synthesizing an inactivating prefix before synthesizing the active portion of the protein molecule, premature physiological function is avoided.

The position of zymogen activation in the overall scheme of physiological control processes is diagrammatically shown in Fig. 1. The term *zymogen* is being used herein to denote in general an inactive precursor that can be converted to an active protein by the cleavage of one or more peptide bonds. This process is essentially irreversible because, in common with many other hydrolytic reactions, proteolysis is an exergonic reaction under normal physiological conditions and there are no simple biological mechanisms to repair a broken peptide bond. In this respect, zymogen activation differs in kind from the freely reversible mechanisms of allosteric transition or covalent modification (14). Whereas the latter are suited to maintain or modulate a steady state of intermediary metabolites, zymogen activation, by virtue of its operational irreversibility, can effect unidirectional changes in the cellular environment and can induce new physiological functions. This type of initiation is more rapid than that regulated by the selective transcription of a genome and is triggered by signals that operate entirely on the post-translational level. Typical zymogen activation reactions are summarized in Table 1.

In some of these processes, the zymogen is converted to the active protein in a single step, whereas in others the process involves consecutive steps or cascades (2) which serve to amplify small stimuli to major physiological responses. Many zymogen activation reactions may have remained undetected thus far because the precursor becomes activated prior to isolation. Indeed, isolation procedures are usually designed for maximum yield of active protein rather than of zymogen and thus may contravene the demonstration of a zymogen precursor.

* By invitation. From time to time, reviews on scientific and technological matters of broad interest are published in the PROCEEDINGS.

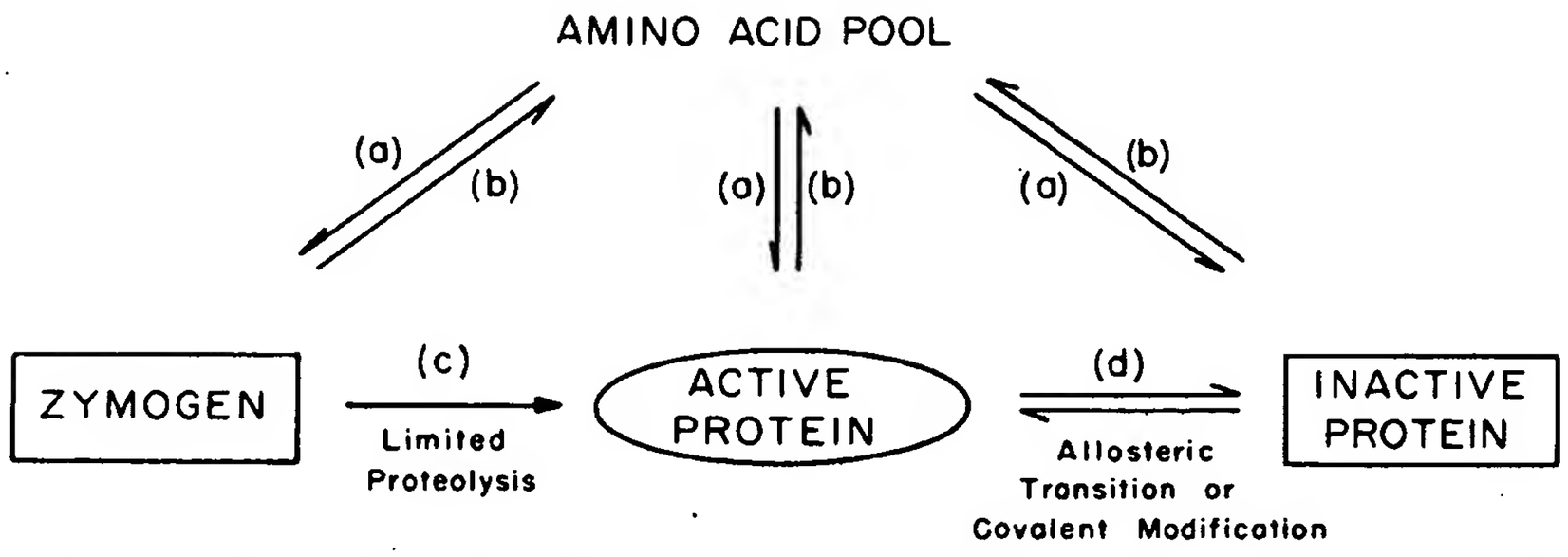


FIG. 1. Schematic representation of major control mechanisms. (a) Transcription and translation regulate the rate of formation of the various proteins from the amino acid pool. (b) Other controls regulate the rate of degradation of the various proteins to their constitutent amino acids. (c) The activity of the protease that catalyzes zymogen activation may be in turn regulated by a series of consecutive reactions of limited proteolysis (Fig. 2). Activation of a zymogen is essentially irreversible *in vivo*. (d) Reversible conformational changes are responsive to effector concentrations or to the activities of specific group transferases and hydrolases.

CONSECUTIVE ZYMOGEN ACTIVATION REACTIONS

A series of consecutive zymogen activation reactions is shown diagrammatically in Fig. 2. X, Y, and Z are zymogens, each having the potential of being converted to an active protein. Conversion of the zymogen X to the protease X_a is triggered by a specific physiological stimulus; in the ensuing cascade, the product of one reaction is a catalyst for the next. The sequence of the events is determined by the specificity of each enzyme and the degree of amplification of the initial stimulus is determined by the efficiency of each activating step. For instance, one molecule of X_a might produce 10^3 molecules of Y_a, which in turn produce 10^6 molecules of active protein. When the zymogen is produced by one cell type and the activating protease by another, communication between the two cell types adds another element to the control mechanism (15). For instance, the activation of pancreatic trypsinogen, which originates in the acinar cells of the pancreas, is triggered by enterokinase, which is secreted from the brush border of the small intestine (16). It should be noted, parenthetically, that the activation of trypsinogen by enterokinase is so specific that only a single bond out of 228 in trypsinogen is cleaved and no other protein has yet been reported to be a substrate for enterokinase (16). Thus, the site of generation of active trypsin is restricted to the confluence of these two secretory streams (Fig. 3). Active trypsin in turn catalyzes the conversion of other pancreatic zymogens to their active forms, i.e., the chymotrypsinogens, proelastase, the procarboxypeptidases, and prophospholipase. This system constitutes a two-stage cascade.

A more complex and extensive cascade system is found in the blood coagulation process, shown in Fig. 4. In fact, the term *cascade* or waterfall was introduced by Macfarlane (17) and by Davie and Ratnoff in 1964 (18) in connection with this series of reactions. Five known proteolytic reactions occur along the so-called intrinsic pathway, which is mediated entirely by components found in the plasma, and four proteolytic steps accompany the extrinsic pathway, which includes factors found in tissues (3). The intrinsic and the extrinsic pathways converge to produce thrombin, which in turn converts fibrinogen to fibrin.

Table 1. Examples of zymogens that are converted to active proteins in response to various stimuli

Normal physiological response	Response to foreign stimulus	Programmed response in development or repair
Vasoactive products	Digestion	Development
Angiotensinogen (61)	Pepsinogen	Prochitin synthetase (7)
Prekallikrein*	Prochymosin	Prococoonase (30)
Kininogen	Trypsinogen*	Procollagenase (68)
Hormones	Chymotrypsinogen	Proacrosin (31)
Proinsulin	Procarboxypeptidases	Fibrinolysis
Proglucagon (62)	Prophospholipase (67)	Plasminogen proactivator*
Proparathyroid hormone	Proelastase	Plasminogen
Enzymes	Blood coagulation	Toxin
Protyrosinase (63, 64)	Factors VII*, IX*, X*	Promellitin (69)
Prephenoloxidase (65)	Prothrombin*	Self-assembly
Prorenin (66)	Factor XIII	Procollagen
	Fibrinogen	Phage head proteins
	Complement	Processing of zymogen precursors (34–37)
	Properdin precursor	
	Factors C_3, C_5*	

Except where indicated by a specific reference, references to these zymogens are found in any of three general sources: *Proteolytic Enzymes*, Methods in Enzymology, Vol. 19 (77); *Peptide Bond Hydrolysis*, The Enzymes, Vol. 3 (78); and *Proteases and Biological Control*, Cold Spring Harbor Conferences on Cell Proliferation, Vol. 2 (79).
* An intermediate component in a series of consecutive zymogen activations.

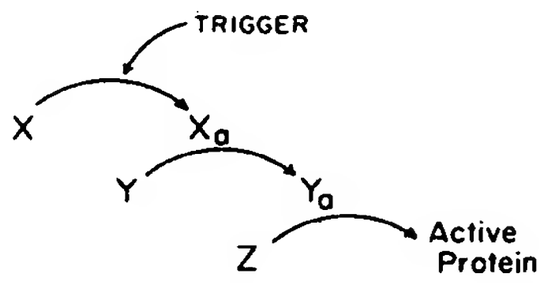


FIG. 2. Schematic representation of consecutive zymogen activation reactions (cascade). Activation of the zymogens X, Y, and Z leads to the formation of the proteases X_a and Y_a and the active protein, in response to a trigger which initiates the conversion of X to X_a.

In addition to the multiple activation steps shown in Fig. 4, processes such as blood coagulation appear to be regulated by specific protein inhibitors (19), which terminate the activation reactions once the product has been generated and restrict the site of blood coagulation to the area of the injury. In this case, a pulse of protease is generated only in the brief interval between zymogen activation and the subsequent formation of an inactive enzyme–inhibitor complex, as represented schematically in Fig. 5. This general scheme illustrates that the physiological signal is transduced to a chemical event by conversion of zymogen X to protease X_a, which in turn converts Y to Y_a and is then inactivated by the inhibitor I_X. In turn, Y_a exists long enough to activate Z and is then inhibited by I_Y. Thus at each step, the duration and breadth of the cascade is controlled to some extent by the concentration of inhibitors, whereas the degree of amplification is determined by the number of steps between X_a and Z. The ultimate change in microenvironment is determined by the nature of the active protein that is finally expressed. There are only a few examples of cascading systems of this kind, but certainly now that the principle is established many more will be found. For instance, premature activation of the pancreatic enzymes shown in Fig. 3 is inhibited at the level of protease Y (trypsin) by pancreatic trypsin inhibitors. The various intermediate proteases of the blood coagulation system are inactivated by specific inhibitors, such as antithrombin III and α_2-macroglobulin, whereas fibrinolysis (i.e., the dissolution of the fibrin clot) and the complement reaction each involve different sets of proteolytic activation reactions and different inhibitors such as α_1-antitrypsin, α_2-macroglobulin, C_1-inactivator, and others (20, 21).

It seems important at this juncture to emphasize the fundamental difference between zymogens on the one hand and enzyme–inhibitor complexes on the other. Both appear inert by activity assays and both can be converted to the active enzyme under the appropriate circumstances. However, as indicated schematically in Fig. 6, zymogens are inactive because the activation peptide confers an altered geometry on the molecule. The structure of the activation peptide is unrelated to that of the active site of the enzyme. On the other hand, macromolecular inhibitors contain regions that are complementary in shape to the active site of the enzyme and some have the characteristics of a pseudosubstrate (22, 23). In fact, only a few residues of the inhibitor molecule interact with the active site of the enzyme, whereas the bulk of the inhibitor serves mainly the function of a supporting structure. This is true of the pancreatic and soybean trypsin inhibitors, as well as of the carboxypeptidase inhibitor from potatoes that has been recently investigated in our laboratory (24).

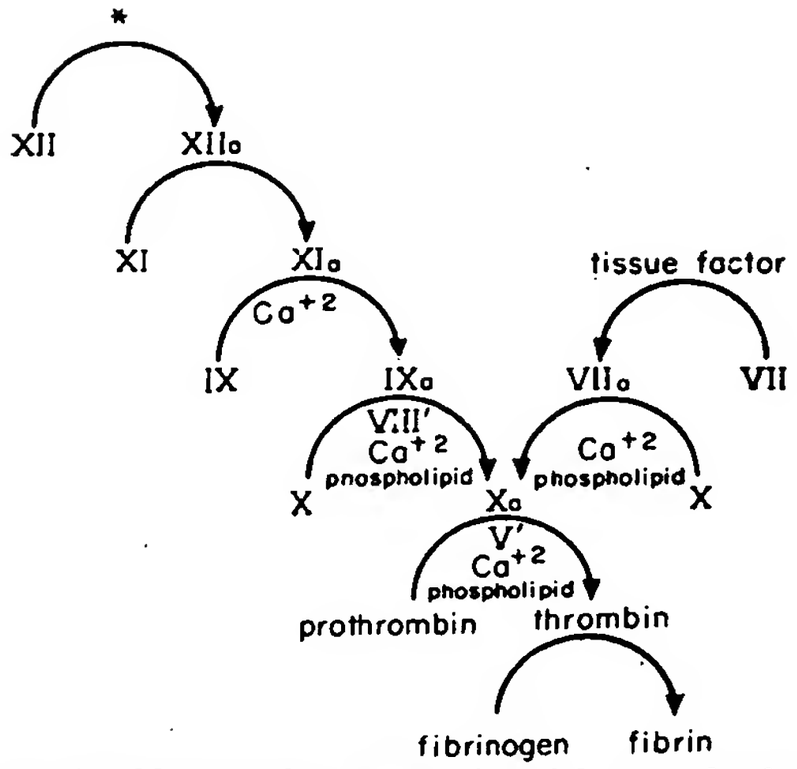


FIG. 4. An abbreviated representation of the cascade of consecutive zymogen activations involved in blood coagulation (3). The intrinsic pathway (left) and the extrinsic pathway (right) converge in activating factor X. The chemical nature of the initial stimulus (*) is not completely understood.

Cascades of zymogen activation reactions do not necessarily operate in isolation but may influence one another in the initial stages, by positive or negative feedback regulation, thus adding another element of control. For instance, three plasma activation systems, the coagulation system, the fibrinolysis system, and the kallikrein system, interact with one another at key points as shown diagrammatically in Fig. 7. The activation of Hageman factor (factor XII) is enhanced by kallikrein, which in turn is generated from prekallikrein under the influence of activated Hageman factor (3, 25, 26). Kallikrein in turn cata-

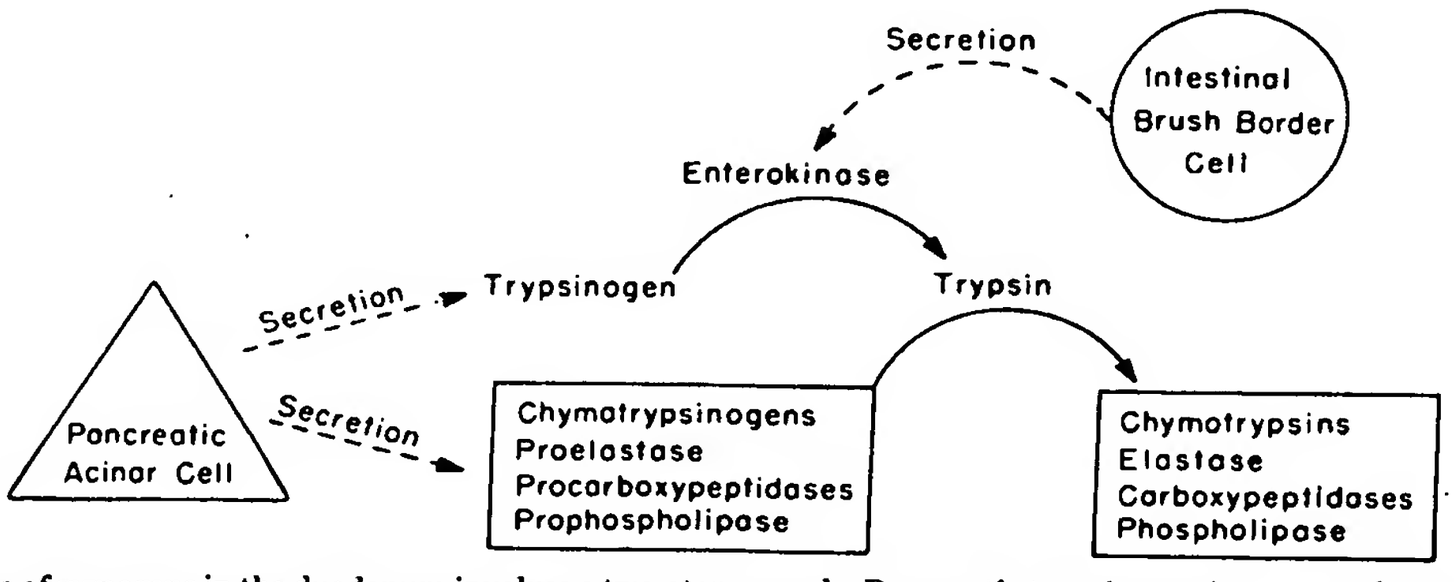


FIG. 3. Activation of zymogens in the duodenum involves a two-step cascade. Pancreatic trypsinogen is converted to trypsin by enterokinase. Trypsin in turn converts the other zymogens to active enzymes.

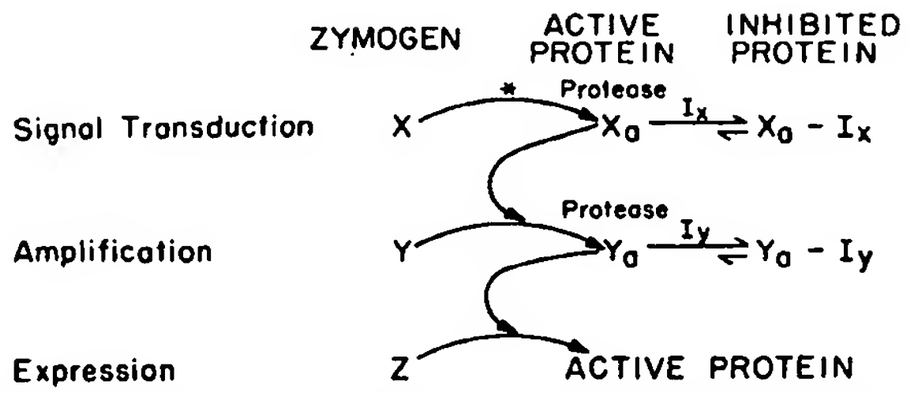


FIG. 5. Generalized scheme of control processes regulated by zymogens, proteases, and their inhibitors. Zymogens (X, Y, and Z) are activated sequentially in processes such as blood coagulation or complement activation. Protease intermediates (X_a, Y_a) may be inactivated by specific inhibitors (I_x, I_y) to limit their action. A physiological signal (*) initiates the cascade by converting zymogen X to protease X_a.

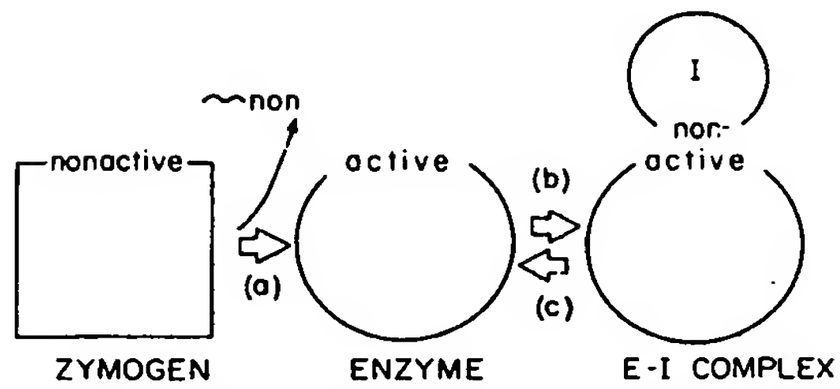


FIG. 6. Schematic illustration of the control of enzyme activity by (a) zymogen activation, (b) formation of an enzyme–inhibitor complex, and (c) dissociation of that complex. ~non represents the activation peptide.

lyzes the formation of kinins (e.g., bradykinin) from their respective precursors, the kininogens (27). Activated Hageman factor is also believed to enhance the conversion of plasminogen proactivator to the activated form, though this process may involve a number of as yet unrecognized intermediates (26). Another focus of interaction of these systems is the degradation of fibrin by plasmin. Some of these reactions can also be inhibited by intermediates of the activation reactions as well as by specific plasma protease inhibitors (28, 29).

At present, descriptions of cascade processes deal primarily with plasma proteases, but fragmentary evidence is consistent with the occurrence of such complex processes elsewhere. For example, it is not yet clear whether the proteases that activate procollagen (8), proinsulin (5), prococoonase (30), and proacrosin (31) are directly triggered by specific enzymes or whether consecutive protease-mediated steps control these transformations. The sequence of discoveries in this field is the reverse of the physiological sequence of events because the starting point for investigation is usually the final product. Subsequent discovery of a precursor of this product leads to inquiries about a catalyst for this transformation; this in turn leads to questions about the origin of this catalyst, and so on. Ultimately one must identify for each process the mechanism that triggers the first step (compare Table 1).

Experimental investigations have usually focused on the physiologically active protein generated by zymogen activation (e.g., an enzyme, hormone, or self-assembling protein), whereas the peptide fragment has been generally ignored. However, recent studies of the activation of complement factor C_3 indicate (32) that a physiologically active peptide, anaphylatoxin (C_{3a}) is released from C_3 together with an active protease (C_{3b}). Analogously, complement factor C_5 and prothrombin each yield upon activation both an active protease and a pharmacologically active peptide (15, 32, 33). It may be suggested that other activation processes also produce dual physiological products in response to a single stimulus and that such reactions serve to express multiple messages in a concerted manner.

Activation reactions involving limited proteolysis have received added significance from the recent observations that secretory proteins, such as the pancreatic zymogens, contain an additional polypeptide extension on the amino side of the normal activation peptide (34–37). Semantically, these proteins are denoted by the prefix *pre-*, whereas zymogens are usually denoted by the prefix *pro-* or the suffix *-ogen* (e.g., *prepro*-parathyroid hormone, *pre*trypsinogen). These "prezymogens" are thought to be the first products of messenger translation and evidence indicates that after passage of the nascent proteins through the membrane of the endoplasmic reticulum the peptide extension is severed from the zymogen by limited proteolysis (34). While the precise significance of these events is currently under investigation, it seems certain that limited proteolysis, analogous to zymogen activation, transforms one precursor to another and that the mechanisms of conversion of zymogens to enzymes may serve as models for the transformation of prezymogen to zymogen.

EVOLUTION OF ZYMOGEN ACTIVATION

If limited proteolysis is indeed a significant physiological control mechanism, how did it come about during biological evolution? Did the primordial cells contain zymogens that subsequently became enzymes, or did they contain enzymes that later became inactivated by extension of the polypeptide chains so as to form zymogens? The available evidence seems to favor, but not prove, the latter alternative, which is also consistent with the empirical observation that relatively few zymogens have been found in prokaryotes or in less highly differentiated eukaryotic organisms.

It is a well-recognized fact that enzymes such as the mammalian serine proteases are a homologous set of proteins which operate by analogous catalytic mechanisms (38, 39). During evolution, those amino acid residues have been preserved that are essential for function, for the maintenance of conformation, or for both. In enzymes, these regions include, in particular, sequences around the active site, and in zymogens they might be expected to include regions around the site of activation. In the case of the serine proteases, the catalytic apparatus includes the so-called "charge-relay" system (40) formed by interaction of amino acid residues which are widely separated from each other in the linear sequence of the molecule. These include, in particular, aspartic acid 102, histidine 57, and serine 195. The alpha amino group of isoleucine 16 forms a salt linkage with aspartic acid 194. (The numbering system is that of bovine chymotrypsinogen A.) An abbreviated presentation of such patterns of homology among serine proteases is given in Fig. 8. The enzymes include three plasma proteases and their homology with the pancreatic serine proteases suggests that both families of enzymes have diverged from a common ancestor (41, 42).

Comparison of the zymogens of these proteases, on first sight, indicates that they are so different in size, composition, and sequence as to preclude a common ancestral relationship. For instance, in the case of bovine trypsinogen, the activation peptide is composed of six residues, whereas in the other extreme, the fragment released during the activation of prothrombin is larger than the enzyme thrombin itself. However, upon closer inspection, at least three classes of activation peptides seem evident. As shown in Fig. 9, trypsinogen activation peptides are homologous and related to the activation peptide released during the generation of the moth enzyme cocoonase from its precursor. Analogously, the activation peptides of

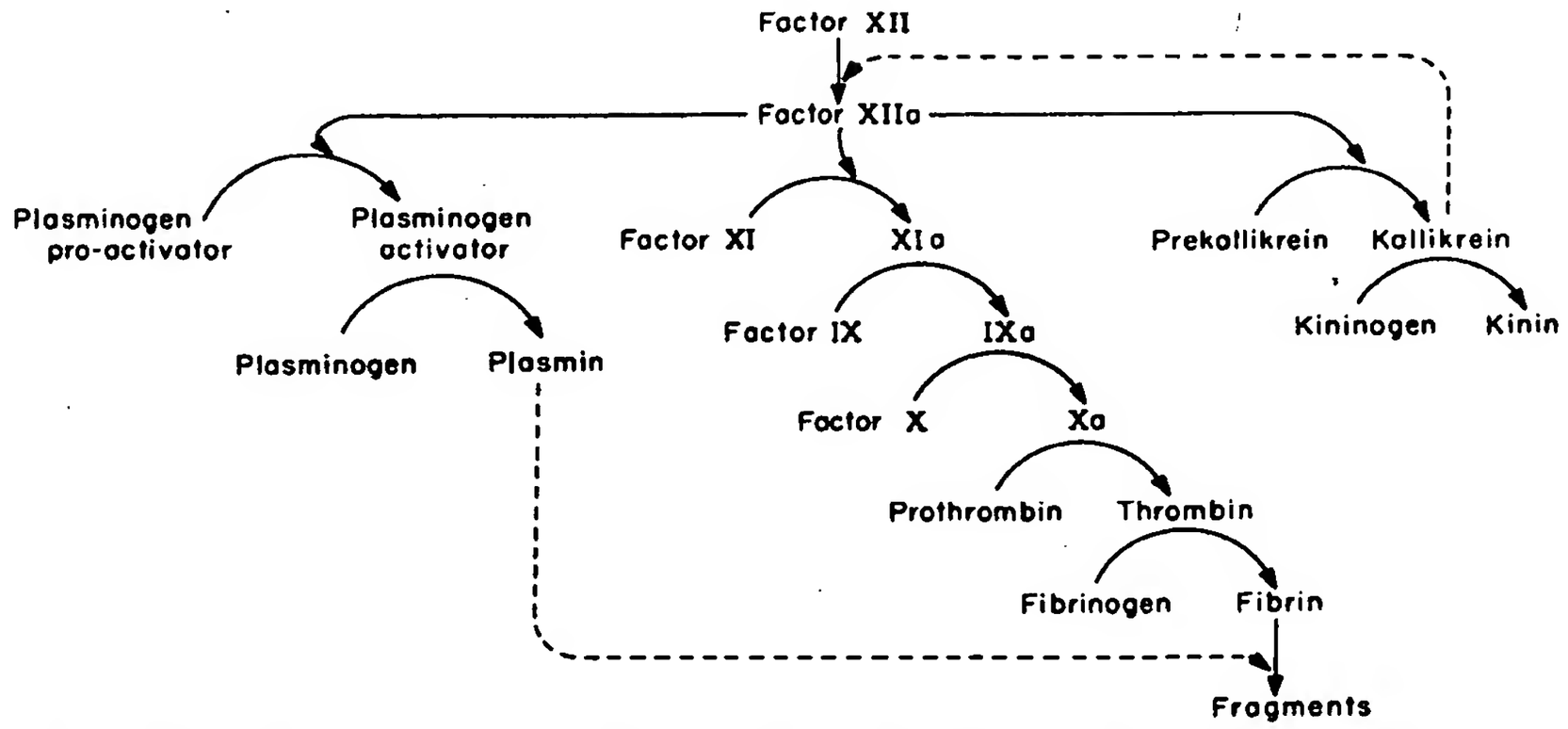


FIG. 7. Interactions among some plasma zymogens and proteases in the fibrinolytic, coagulation, and kinin systems (7, 25–27). For details, see the *text*.

various chymotrypsinogens (A, B, and C) and proelastase are homologous, although they appear to be quite different from those of the trypsinogen group. A third structural class of activation peptides comprises certain blood coagulation proteases, particularly those of prothrombin, factor X, and factor IX. The structural relationship among these zymogens is shown schematically in Fig. 10. Whereas prothrombin and factor IX are synthesized as single polypeptide chains and only subsequently during activation are converted into two-chain structures, factor X appears to consist even initially of two chains (7). The amino-terminal sequences of the single chains of prothrombin and factor IX and of the light chain of factor X show evidence of homology, but in the case of prothrombin and factor X no such similarity is evident in the sequences immediately preceding the peptide bond that is cleaved during the activation reaction (43, 44). If in fact the activation peptides of the three groups of serine proteases, i.e., the trypsinogens, the chymotrypsinogens, and the blood coagulation proteases, are unrelated, one would conclude that the homologous enzymes arose first and that the prefixes were subsequently added as independent events following gene duplication and divergence of functions (45). However, one cannot rule out the possibility that the activation peptides of the serine proteases represent distant homologous sequences which are the products of a rapidly mutating gene, as for instance in the case of the fibrinopeptides. de Haen *et al.* (46) have in fact argued for such a homologous relationship among these various zymogens.

The activation peptides of the three blood coagulation zymogens are also of interest because they illustrate that the removal of the activation peptide during zymogen activation may have additional consequences besides the generation of an active site. It has been recently demonstrated by several workers that the amino-terminal region of prothrombin, and probably also of factors X and IX, contains an unusual amino acid, γ-carboxyglutamic acid, which appears to form an effective site for chelating calcium ions (43, 47–49). Formation of this amino

	16	57	102	195
PANCREATIC PROTEASES				
Bovine chymotrypsin A	Ile-Val-Asn-Gly-----	Ala-Ala-His-Cys-----	Asn-Asp-Ile-----	Gly-Asp-Ser-Gly-
Bovine trypsin	Ile-Val-Gly-Gly-	-Ala-Ala-His-Cys-	-Asn-Asp-Ile-	-Gly-Asp-Ser-Gly-
Porcine trypsin	Ile-Val-Gly-Gly-	-Ala-Ala-His-Cys-	-Asn-Asp-Ile-	-Gly-Asp-Ser-Gly-
Dogfish trypsin	Ile-Val-Gly-Gly-	-Ala-Ala-His-Cys-	-Asn-Asp-Ile-	-Gly-Asp-Ser-Gly-
Porcine elastase	Val-Val-Gly-Gly-	-Ala-Ala-His-Cys-	-Tyr-Asp-Ile-	-Gly-Asp-Ser-Gly-
PLASMA PROTEASES (Bovine)				
Thrombin	Ile-Val-Glu-Gly-	-Ala-Ala-His-Cys-	-Arg-Asp-Ile-	-Gly-Asp-Ser-Gly-
Factor Xa	Ile-Val-Gly-Gly-	-Ala-Ala-His-Cys-	-Phe-Asp-Ile-	-Gly-Asp-Ser-Gly-
Factor IXa	Val-Val-Gly-Gly-	----- n.d. ------	---- n.d.----	-Gly-Asp-Ser-Gly-
Plasmin (human)	Val-Val-Gly-Gly-	-Ala-Ala-His-Cys-	---- n.d.----	-Gly-Asp-Ser-Gly-
OTHER PROTEASES				
Cocoonase	Ile-Val-Gly-Gly-	----- n.d. ------	---- n.d.----	-Gly-Asp-Ser-Gly-

FIG. 8. Similarities in amino acid sequences immediately adjacent to components of the active sites of various serine proteases (43, 44, 48, 70–75). The residue numbers are those of chymotrypsinogen A. Underlined residues differ from the majority shown for a particular position in the sequence. n.d. = not determined.

GROUP I			
Bovine trypsinogen		Val-Asp-Asp-Asp-Asp-Lys	Ile-
Dogfish trypsinogen		Ala-Pro-Asp-Asp-Asp-Asp-Lys	Ile-
Prococoonase		(5 residues)---Arg-Thr-Gln-Asp-Asp-Gly-Gly-Lys	Ile-
GROUP II			
Bovine chymotrypsinogen A		(7 res)---Pro-Val-Leu-Ser-Gly-Leu-Ser-Arg	Ile-
Bovine chymotrypsinogen B		(7 res)---Pro-Val-Leu-Ser-Gly-Leu-Ala-Arg	Ile-
Bovine chymotrypsinogen C		(5 res)---Phe-Gln-Pro-Asn-Leu-Ser-Ala-Arg	Val-
Lungfish proelastase A		(3 res)---Pro-Ser-Tyr-Pro-Pro-Thr-Ala-Arg	Val-
GROUP III			
Bovine prothrombin	Ala-Asn-Lys-Gly-Phe-Leu-Gla-Gla------(307 res)	---------Phe-Glu-Ser-Tyr-Ile-Glu-Gly-Arg	Ile-
Bovine Factor X	Ala-Asn-Ser- - -Phe-Leu-Gla-Gla--(133 res)	(43 res)--Glu-Asp-Gly-Ser-Gln-Val-Val-Arg	Ile-
Bovine Factor IX	Tyr-Asn-Ser-Gly-Lys-Leu-Glx-Glx----(ca. 147 res)	----------	Val-
Bovine plasminogen	Glu-Pro-Leu-Asp-Asp-Tyr-Val-Asn----(ca. 550 res)	----------Arg	Val-

FIG. 9. Homologous amino-terminal sequences of three groups of zymogens of serine proteases (30, 41, 43, 44, 46, 48, 70, 76). The vertical arrow denotes the site of limited proteolysis during zymogen activation. Gla indicates γ-carboxyglutamic acid.

acid is a prerequisite for the activation of prothrombin by factor X, and its absence in certain patients is associated with disorders of the blood coagulation process (49). The activation of prothrombin by factor X_a is accelerated by calcium ions, as is the activation of factor X by factor IX_a. Both activation reactions require, in addition, phospholipid and other coagulation proteins (factors V and VIII, respectively). Whereas activated factor X (X_a) stays bound to the phospholipid micelle, thrombin does not, because the calcium-binding site is released during the activation of prothrombin, but the binding site of factor X remains attached to this zymogen even after activation (3).

MECHANISM OF ZYMOGEN ACTIVATION

It is of foremost interest to inquire about the mechanism by which the cleavage of a single polypeptide bond in a zymogen generates enzymatic function. Does limited proteolysis merely remove an obstruction from the active site, or does it induce a conformational change that generates an active site? It is a well-accepted dogma that the amino acid sequence of a protein provides the determinants of its three-dimensional structure and that chemical alterations of the backbone or the side chains could disrupt the balance of forces that dictate a particular conformation. A change in conformation could induce the *de novo* assembly of the catalytic apparatus, the generation of the substrate-binding site, or the removal of an obstruction from both (see Fig. 11). In the best-studied cases, i.e., the x-ray crystallographic comparison of the structures of chymotrypsinogen and chymotrypsin, the observed differences were so subtle that it was not possible to decide unambiguously which structural changes were crucial for zymogen activation and which were incidental (50). The recent discovery that many zymogens possess weak but intrinsic enzymatic activity has made it possible to examine the activation process by kinetic

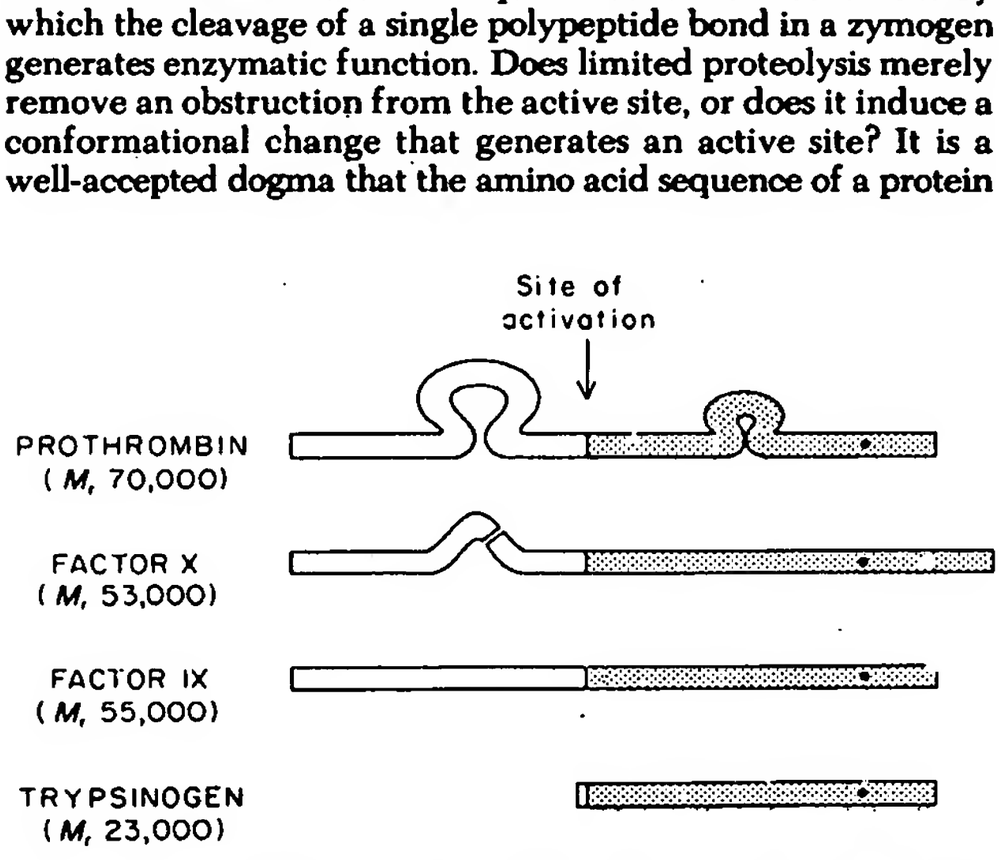


FIG. 10. Comparison of the lengths of the polypeptide chains of the zymogen forms of four homologous serine proteases. The stippled segments denote the homologous enzymes and the dots the reactive serines. M_r = molecular weight.

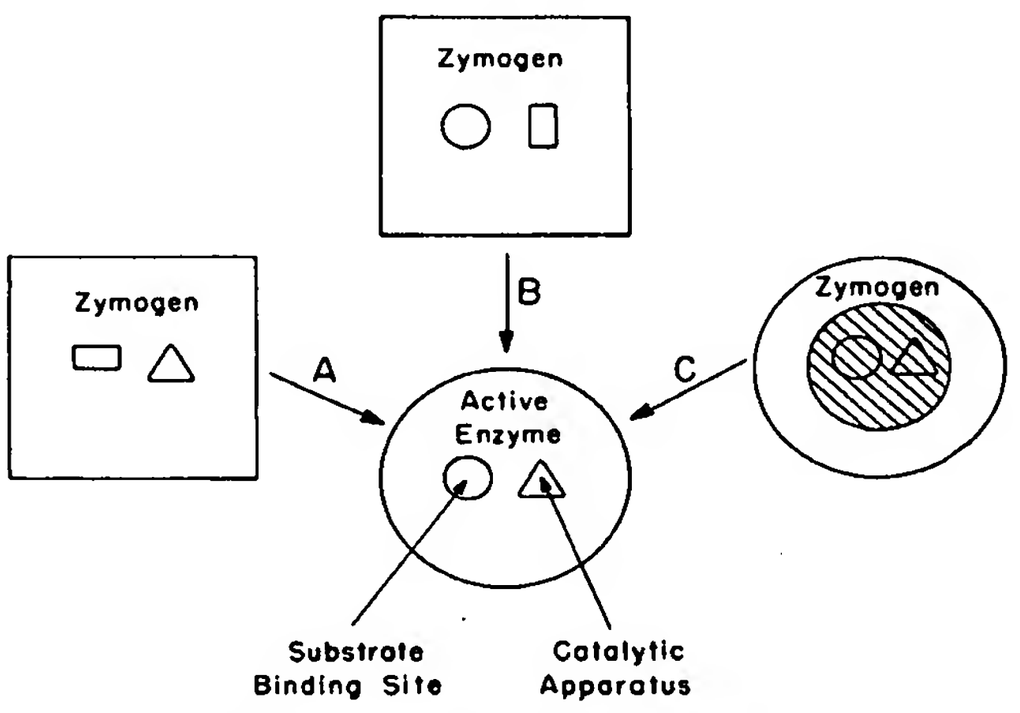


FIG. 11. Schematic illustration of three possible modes of zymogen activation. Pathway A: the conversion of zymogen to enzyme induces a conformational change that improves the substrate binding site. Pathway B: the binding site preexists in the zymogen and zymogen activation induces the formation of an effective catalytic site. Pathway C: the activation peptide (hatched) occludes in the zymogen the entire active site, which becomes exposed only after removal or dislocation of this peptide segment.

and spectral analysis of the zymogen before activation and of the enzyme afterwards (51–55). These studies, to be described briefly in the following paragraph, indicate that the catalytic apparatus of trypsin, chymotrypsin, and probably other zymogens of serine proteases, is largely preexistent in the zymogen forms and that during activation the effectiveness of the binding site is improved more than 1000-fold. The evidence can be briefly summarized as follows.

Several zymogens, like their respective enzymes, react stoichiometrically with active-site-directed inhibitors such as diisopropylphosphofluoridate to form stable covalent compounds at a reaction rate which is 10^4 to 10^5 times slower than that of enzyme (52). Certain other pseudo-substrates react not only with the enzyme but also with the parent zymogen. In both cases stable acyl-intermediates can be isolated at low pH and then deacylated at higher pH (54, 55). The second-order rate constants of acylation are 10^4 to 10^5 times lower for the zymogen than for the enzyme, but the deacylation rates differ only by factors ranging from 2 to 70. The pseudo-substrate methane sulfonyl fluoride, which has a low affinity for the enzyme, presumably because it is not bound to the substrate-binding pocket, also has a low affinity for the zymogen but the second-order rate constant differs not by 10^4 or 10^5 as in the case of diisopropylphosphofluoridate, but by only about 50-fold (56). Finally, a competitive inhibitor (*p*-aminobenzamidine) is bound by trypsinogen 10^3 to 10^4 times less firmly than by trypsin, presumably because the substrate-binding pocket is not fully developed prior to activation (54). All of these observations agree with the idea that in these zymogens the substrate-binding pocket is distorted but that the catalytic apparatus is largely preexistent.

It should be recognized, however, that the effectiveness of enzymes is determined by the geometries of both the enzyme and the substrate and that these have to be compatible to form the productive transition state during catalysis [in the case of serine proteases this appears to be a tetrahedral intermediate (57, 58)]. With ideal substrates, the geometries of enzyme and substrate are suitable to meet these requirements, whereas with poor substrates and pseudosubstrates the complex is less productive because of misalignment of susceptible bonds relative to the idealized state (M. A. Kerr, K. A. Walsh, and H. Neurath, manuscript in preparation). Zymogens may be relatively poor catalysts for a certain reaction because the geometry of the active site is unfavorable to bind substrates in a productive mode and to form the idealized transition state. Chymotrypsinogen and trypsinogen are inferior catalysts for all known substrates but activation by limited proteolysis changes the conformation of these zymogens, improves the binding of specific substrates, and also allows the transition state to be reached.

In a few cases, however, such conformational transitions occur seemingly without peptide bond cleavage. For instance, the conversion of human plasminogen to plasmin is induced by streptokinase, and as a first step the formation of a complex of streptokinase and plasminogen alone suffices to induce enzymatic activity without peptide bond cleavage (59, 60). Similarly, the activation of factor XII (Hageman factor) and of complement factor C_1 does not appear to involve peptide bond cleavage (4, 26). In the case of factor X, peptide bond cleavage is required for activation but the differences in enzymatic activity toward pseudosubstrates between zymogen and active enzyme are not nearly as great as in the case of the model system, chymotrypsinogen and chymotrypsin (M. A. Kerr, K. A. Walsh, and H. Neurath, manuscript in preparation). It thus appears that the various zymogens and their enzymes differ from each in their relative catalytic efficiency, depending upon the magnitude of the conformational changes required for binding and aligning the substrate in an optimally productive mode.

CONCLUSIONS

While the activations of pancreatic zymogens such as trypsinogen and chymotrypsinogen are the best understood examples of the induction of biological activity by limited proteolysis, it is important to emphasize that they represent only examples of a great variety of physiological functions that are induced by limited peptide bond cleavage of precursor proteins. The product of activation may be an enzyme, a hormone, a pharmacologically active peptide, or a structural component of tissues, but the ultimate chemical event is limited proteolysis in every instance. Proteases can thus generate functions but they can also destroy them, and in this sense it is useful to think of limited proteolysis as a control element that can turn other reactions on and off by generating or destroying their catalysts. The on and off reactions are controlled by different switches, so to speak, because proteolysis is essentially an irreversible process and proteases are not endowed with repair functions. Zymogens are poised to respond to signals, to amplify them, and to respond to them irreversibly. Examples of such control processes are being increasingly observed and should be further thought of in the exploration of the control of metabolism and other physiological phenomena, including processes of development and differentiation.

This work has been supported in part by grants from the National Institutes of Health (GM15731) and the American Cancer Society (BC-91R).

1. Neurath, H. (1957) *Adv. Protein Chem.* **12**, 320–386.
2. Neurath, H. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkins, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 51–64.
3. Davie, E. W. & Fujikawa, K. (1975) *Annu. Rev. Biochem.* **44**, 799–829.
4. Müller-Eberhard, H. J. (1975) *Annu. Rev. Biochem.* **44**, 697–724.
5. Steiner, D. F., Kemmler, W., Tager, H. S., Rubenstein, A. H., Lernmark, A. & Zuhlke, H. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 531–549.
6. Zanefeld, L. J. D., Polakoski, K. L. & Schumacher, G. F. B. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 683–706.
7. Cabib, E. & Farkas, V. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 2052–2056.
8. Bornstein, P. (1974) *J. Supramol. Struct.* **2**, 108–120.
9. Laemmli, U. K. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 661–676.
10. Linderstrøm-Lang, K. U. & Ottesen, M. (1949) *C. R. Trav. Lab. Carlsberg* **26**, 403–442.
11. Richards, F. M. & Vithayathil, P. J. (1959) *J. Biol. Chem.* **234**, 1459–1465.
12. Schimke, R. T. & Doyle, D. (1970) *Annu. Rev. Biochem.* **39**, 929–976.
13. Ottesen, M. (1967) *Annu. Rev. Biochem.* **36**, 55–76.
14. Stadtman, E. R. (1970) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), 3rd ed., Vol. 1, pp. 397–459.
15. Walsh, K. A. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 1–11.
16. Maroux, S., Baratti, J. & Desnuelle, P. (1971) *J. Biol. Chem.* **246**, 5031–5039.
17. Macfarlane, R. G. (1964) *Nature* **202**, 498–499.

18. Davie, E. W. & Ratnoff, O. D. (1964) *Science* **145**, 1310–1312.
19. Heimburger, N. (1974) in *Bayer-Symposium V, Proteinase Inhibitors*, eds. Fritz, H., Tschesche, H., Greene, L. J. & Truscheit, E. (Springer-Verlag, New York), pp. 14–22.
20. Harpel, P. C., Mosesson, M. W. & Cooper, N. R. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 387–404.
21. Jeppsson, J. and Laurell, C. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 405–414.
22. Huber, R., Kikla; D., Steigemann, W., Deisenhofer, J. & Jones, A. in *Bayer Symposium V, Proteinase Inhibitors*, eds. Fritz, H., Tschesche, H., Green, L. J. & Truscheit, E. (Springer-Verlag, New York), pp. 497–512.
23. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974) *Biochemistry* **13**, 4212–4228.
24. Hass, G. M., Ako, H., Grahn, D. T. & Neurath, H. (1976) *Biochemistry* **15**, 93–100.
25. Nagasawa, S., Takahashi, H., Koida, M., Suzuki, T. & Schoenmakers, J. G. G. (1968) *Biochem. Biophys. Res. Commun.* **32**, 644–649.
26. Ulevitch, R. J., Cochrane, C. G., Revak, S. D., Morrison, D. C. & Johnston, A. R. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 85–93.
27. Pisano, J. J. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 199–222.
28. Astrup, T. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 343–355.
29. Dan, K. & Reich, E. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 357–366.
30. Felsted, R. L., Kramer, K. J., Law, J. H., Berger, E. & Kafatos, F. C. (1973) *J. Biol. Chem.* **248**, 3012–3020.
31. Meizel, S. & Mukerji, S. K. (1975) *Biol. Reprod.* **13**, 83–93.
32. Hugli, T. E. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 273–290.
33. Jesty, J. & Esnouf, M. P. (1973) *Biochem. J.* **131**, 791–799.
34. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–862.
35. Devillers-Thiery, A., Kindt, T., Scheele, G. & Blobel, G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 5016–5020.
36. Kemper, B., Habener, J. F., Ernst, M. D., Potts, J. T., Jr. & Rich, A. (1976) *Biochemistry* **15**, 15–19.
37. Schechter, I. & Burstein, Y. (1976) *Biochem. Biophys. Res. Commun.* **68**, 489–496.
38. Neurath, H., Walsh, K. A. & Winter, W. P. (1967) *Science* **158**, 1638–1644.
39. Hartley, B. S. & Shotton, D. M. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), 3rd ed., Vol. 3, pp. 323–373.
40. Blow, D. M. (1976) *Acc. Chem. Res.* **9**, 145.
41. Enfield, D. L., Ericsson, L. H., Fujikawa, K., Titani, K., Walsh, K. A. & Neurath, H. (1974) *FEBS Lett.* **47**, 132–135.
42. Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H. & Davie, E. W. (1972) *Biochemistry* **11**, 4899–4903.
43. Enfield, D. L., Ericsson, L. H., Walsh, K. A., Neurath, H. & Titani, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 16–19.
44. Titani, K., Fujikawa, K., Enfield, D. L., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3082–3086.
45. Neurath, H., Walsh, K. A. & Gertler, A. (1974) in *Metabolic Interconversion of Enzymes, 1973*, eds. Fischer, E. H., Krebs, E. G., Neurath, H. & Stadtman, E. R. (Springer-Verlag, New York), pp. 301–312.
46. de Haen, C., Neurath, H. & Teller, D. C. (1975) *J. Mol. Biol.* **92**, 225–259.
47. Stenflo, J. (1974) *J. Biol. Chem.* **249**, 5527–5535.
48. Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Morris, H. R. & Dell, A. (1974) *FEBS Lett.* **44**, 189–193.
49. Stenflo, J., Fernlund, P. & Roepstorff, P. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 111–122.
50. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. & Xuong, N. H. (1970) *Biochemistry* **9**, 1997–2009.
51. Robinson, N. C., Neurath, H. & Walsh, K. A. (1973) *Biochemistry* **12**, 420–426.
52. Morgan, P. H., Robinson, N. C., Walsh, K. A. & Neurath, H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3312–3316.
53. Kassell, B. & Kay, J. (1973) *Science* **180**, 1022–1027.
54. Gertler, A., Walsh, K. A. & Neurath, H. (1974) *Biochemistry* **13**, 1302–1310.
55. Kerr, M. A., Walsh, K. A. & Neurath, H. (1975) *Biochemistry* **14**, 5088–5094.
56. Morgan, P. H., Walsh, K. A. & Neurath, H. (1974) *FEBS Lett.* **41**, 108–110.
57. Henderson, R., Wright, C. S., Hess, G. P. & Blow, D. M. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 63–70.
58. Robertus, J. D., Kraut, J., Alden, R. A. & Birktoft, J. J. (1972) *Biochemistry* **11**, 4293–4303.
59. Nagendra, K., Reddy, N. & Markus, G. (1972) *J. Biol. Chem.* **247**, 1683–1691.
60. Summaria, L., Arzadon, L., Bernabe, P. & Robbins, K. C. (1974) *J. Biol. Chem.* **249**, 4760–4769.
61. Skeggs, L. T., Jr., Kahn, J. R., Lentz, K. & Shumway, N. P. (1957) *J. Exp. Med.* **106**, 439–453.
62. Tager, H. S. & Steiner, D. F. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2321–2325.
63. Holstein, T. J., Stowell, C. P., Quevedo, W. C., Jr., Zarcaro, R. M. & Bienieki, T. C. (1973) *Yale J. Biol. Med.* **46**, 560–571.
64. McGuire, J., Newman, D. & Barisas, G. (1973) *Yale J. Biol. Med.* **46**, 572–582.
65. Dohke, K. (1973) *Arch. Biochem. Biophys.* **157**, 203–209.
66. Leckie, B. (1973) *Clin. Sci.* **44**, 301–304.
67. de Haas, G. H., Slotboom, A. J., Bonsen, P. P. M., Van Deenen, L. L. M., Maroux, S., Puigserver, A. & Desnuelle, P. (1970) *Biochim. Biophys. Acta* **221**, 31–53.
68. Harper, E., Bloch, K. J. & Gross, J. (1971) *Biochemistry* **10**, 3035–3041.
69. Kreil, G. (1973) *Eur. J. Biochem.* **33**, 558–566.
70. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C.), Vol. 5.
71. Hermodson, M. A., Ericsson, L. H., Neurath, H. & Walsh, K. A. (1973) *Biochemistry* **12**, 3146–3153.
72. Titani, K., Ericsson, L. H., Neurath, H. & Walsh, K. A. (1975) *Biochemistry* **14**, 1358–1366.
73. Robbins, K. C., Bernabe, P., Arzadon, L. & Summaria, L. (1972) *J. Biol. Chem.* **247**, 6757–6762.
74. Robbins, K. C., Bernabe, P., Arzadon, L. & Summaria, L. (1973) *J. Biol. Chem.* **248**, 1631–1633.
75. Kramer, K. J., Felsted, R. L. & Law, J. H. (1973) *J. Biol. Chem.* **248**, 3021–3028.
76. Wallen, P. & Wiman, B. (1975) in *Proteases and Biological Control*, eds. Reich, E., Rifkin, D. B. & Shaw, E. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2, pp. 291–303.
77. Perlmann, G. E. & Lorand, L., eds. (1970) *Methods in Enzymology* (Academic Press, New York), Vol. 19.
78. Boyer, P. D., ed. (1971) *The Enzymes*, 3rd ed. (Academic Press, New York), Vol. 3.
79. Reich, E., Rifkin, D. & Shaw, E., eds. (1975) *Cold Spring Harbor Conferences on Cell Proliferation* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 2.

Annu. Rev. Genet. 2004. 38:709–24
doi: 10.1146/annurev.genet.38.072902.093416

First published online as a Review in Advance on July 30, 2004

PROTEOLYSIS AS A REGULATORY MECHANISM

Michael Ehrmann[1] and Tim Clausen[2]
[1]*Cardiff University, School of Biosciences, Cardiff CF10 3US, United Kingdom; email: ehrmann@cf.ac.uk*
[2]*Institute for Molecular Pathology (IMP), A-1030 Vienna, Austria; email: tim.clausen@imp.univie.ac.at*

Key Words DegS, YaeL, APP, RIP, HtrA

■ **Abstract** Proteases can play key roles in regulation by controlling the levels of critical components of, for example, signal transduction pathways. Proteolytic processing can remove regulatory proteins when they are not needed, while transforming others from the dormant into the biologically active state. The latter mechanism often involves a subsequent change of cellular localization such as the movement from the membrane to the nucleus. The investigation of these processes has revealed a new type of proteolytic activity, regulated intramembrane proteolysis, and a reversible switch in activity occurring in the HtrA family of serine proteases. The bacterial RseA and the human amyloid precursor processing pathways are used as models to review these novel principles that are evolutionarily conserved and have wide biological implications.

CONTENTS

0066-4197/04/1215-0709$14.00

INTRODUCTION

Proteases, Substrates, and Regulation

Proteases and peptidases refer to the same group of enzymes whose catalytic function is to hydrolyze peptide bonds, leading to the breakdown of substrates. They can be divided into five different groups, depending on the active site residue or ion that carries out catalysis. The molecular mechanisms of the corresponding serine, threonine, cysteine, aspartic, and metallo proteases have been well studied and are for most proteases well understood (for review see 59–63). The physiological roles of proteases are very diverse, ranging from digestive functions, the removal of damaged proteins, and protein maturation to the precise processing of regulatory proteins. The biological implications of proteolytic reactions are so immense that perhaps no other class of proteins is involved in such a diverse array of physiological processes. On the conceptual level, proteases provide an important link between genetics and biochemistry as they often act as key players in regulatory events, for example in important signaling cascades such as apoptosis (18, 67) and the regulation of the unfolded protein response (27, 54). To fully appreciate some recently identified novel principles, classical features are introduced first.

Proteases do not attack their protein substrates at random. Rather, they display a high degree of specificity in identifying (binding) and processing their substrates. Substrate recognition depends on a variety of factors, such as colocalization of protease and substrate in the same cellular compartment and on substrate specificity. Substrate specificity is often defined by the structural properties of the active site or sometimes by so-called adaptor proteins that can feed substrate to the protease (27). Another obvious prerequisite is the surface accessibility of potential cleavage sites. Cleavage sites that are hidden in native proteins can become accessible as a result of unfolding or by loss of an interaction partner. A variation of the surface accessibility concept is the compartmentalization of active sites within a large protease particle. Here, as exemplified by the proteasome or the DegP protease from *Escherichia coli*, the active sites are located in an inner cavity that can be accessed only by unfolded proteins (12, 48). This principle is an elegant way to prevent degradation of folded proteins. Another relevant phenomenon is the sequentiality of degradation events. Proteases involved in regulatory processes, for example, often work in a concerted manner whereby more than one protease is involved in processing of a given substrate.

Whereas substrate accessibility presents one way of regulating protease activity, other general principles include protease activation and inhibition. Many of the classical proteases such as trypsin are synthesized as zymogens, inactive forms that are converted into the active form by proteolytic excision of a propeptide (41, 49). Zymogen activation was long considered to be a rapid and irreversible response to a relevant signal that leads to constitutively active protease. However, recent reports indicate reversible zymogen activation where either other proteins (24) or external factors such as concentrated salts (31) or temperature (69) can switch protease

activity on or off. Yet another level of regulation of protease activity is provided by various natural inhibitors. According to the principle of exosite binding, these inhibitors have in common that they are tightly bound to the protease by occupying the substrate-specificity pockets (for review see 5, 83).

Regulatory Proteolysis

Not only is the activity of proteases subject to fine-tuned regulation but many proteases are themselves part of regulatory mechanisms. Many proteins including receptors, kinases, transcription factors, and structural components become modified by proteolysis in order to gain activity from a latent state or to alter an existing function. Furthermore, proteolysis is the most common mechanism for inactivating proteins. Thus, proteases play important roles by modulating enzyme activity and protein-protein interactions within multicomponent signaling pathways, thereby acting as master regulators. Classical examples include the Lon protease and its role in the regulation of the lytic development of phage lambda (26) and the regulation of the heat shock response in *E. coli* (2). As this emerging topic is of great interest, some phenomena have already been described elsewhere (27, 33). This review focuses on two systems, the RseA/sigma E (σ^E) pathway of the bacterial unfolded protein response (UPR) and the amyloid precursor protein (APP) pathway of higher organisms. These systems were chosen as models to demonstrate a number of key features, classical and newly discovered, in proteolysis as a regulatory mechanism. Furthermore, they share common mechanistic principles, although the players involved are different.

Regulated Intramembrane Proteolysis

Both the RseA/σ^E and the APP pathways apply a novel proteolytic mechanism, termed regulated intramembrane proteolysis (RIP), which has been discovered only recently (for review see 7, 75, 80). RIP proteases are polytopic integral membrane proteins that cleave their substrates within the membrane. They represent a novel family of proteolytic enzymes in which the conserved active site residues are part of transmembrane segments, indicating that a common mechanistic principle is conserved within different active sites. RIP proteases can be monomers or complex heterooligomers and can belong to the aspartic, metallo-, or serine protease clans (21, 76, 79). These proteases are widely conserved and are present in bacteria, archaea, and eukaryotes (80). Many RIP proteases are part of diverse signaling cascades. Generally, their substrates are integral membrane proteins, often with only one transmembrane segment, that are in a resting state before cleavage. Cleaved products are released from the membrane and become biologically active by moving either into the nucleus or to other relevant cellular compartments. Typically, RIP requires initial processing of the substrate. For example, the RseA and APP proteins are truncated by the DegS protease and β-secretase, respectively, before they are further processed via RIP.

THE UNFOLDED PROTEIN RESPONSE REGULATES PROTEIN QUALITY CONTROL IN THE CELL ENVELOPE OF GRAM-NEGATIVE BACTERIA

The cell envelope of gram-negative bacteria such as *E. coli* comprises the periplasm and the outer membrane. The periplasm contains the proteoglycan layer, which provides cell shape and rigidity. It occupies about 30% of the cell volume (82) and has a higher protein concentration compared with the cytoplasm (6). The outer membrane hosts about 350 different proteins, of which most share structural elements and mechanisms of assembly (4, 15, 46, 73). The cell envelope is responsible for communication with the environment (e.g., nutrient uptake, exclusion of toxic compounds, the interaction with host cells, other bacteria, or bacteriophages). These vital functions are mediated by proteins whose folding, functionality, and abundance must be continuously monitored for optimal performance of the entire organism. Given the large number of proteins, their different properties, and abundances, protein quality control, repair, and degradation are challenging tasks. The individual cellular factors and regulatory systems involved in protein quality control that are well studied include the molecular chaperone Skp, four proline isomerases (SurA, FkpA, PpiD, and PpiA), the disulfide redox machinery Dsb, and some of the many peptidases and proteases (for review see 36, 58).

The regulation of protein quality control, repair, and degradation is carried out by various signaling mechanisms that are known as UPR. Three signal transduction pathways σ^E, Cpx, and Bae regulate the cellular response to extracytoplasmic folding stress (for review see 56, 57). It is generally accepted that misfolded proteins, protein fragments, and mislocalized membrane proteins are activators of these stress response pathways. Such activation occurs under any kind of stress that causes protein-folding problems such as heat shock, pH, and osmotic stress, stress caused by the immune system, or when proteins are overproduced at nonnative levels. Therefore, protein quality control systems are essential determinants of cell fate. Even under nonstress conditions these systems play important roles, for example in the normal folding pathways of individual proteins, in detection and "treatment" of aged proteins, the assembly of protein complexes, and the insertion of membrane proteins.

The RseA/σ^E Pathway

The σ^E pathway is essential for viability, and its central component is the alternative sigma factor RpoE (σ^E) (17). As a sigma factor, RpoE interacts with RNA polymerase and confers transcription of specific stress promoters that contain appropriate binding sites (Figure 1*A*). As sensing of folding problems and transcriptional activation of stress promoters occur in different cellular compartments, there are additional factors in the cell envelope that regulate the activity of σ^E: the antisigma factor RseA, the RseB protein that modulates the stability of the σ^E-RseA complex and RseC, whose function remains to be determined.

RseA is an integral cytoplasmic membrane protein that comprises an N-terminal cytoplasmic domain, one transmembrane segment, and a periplasmic domain. The cytoplasmic segment of RseA contains the antisigma function and binds to σ^E, keeping it in the inactive conformation. A recent crystal structure of the cytoplasmic domain of RseA complexed with σ^E supports a 1:1 stoichiometry and shows that the antisigma domain is sandwiched between two σ^E domains. The inhibition of σ^E occurs because the RseA domain sterically blocks binding of σ^E to RNA polymerase (9).

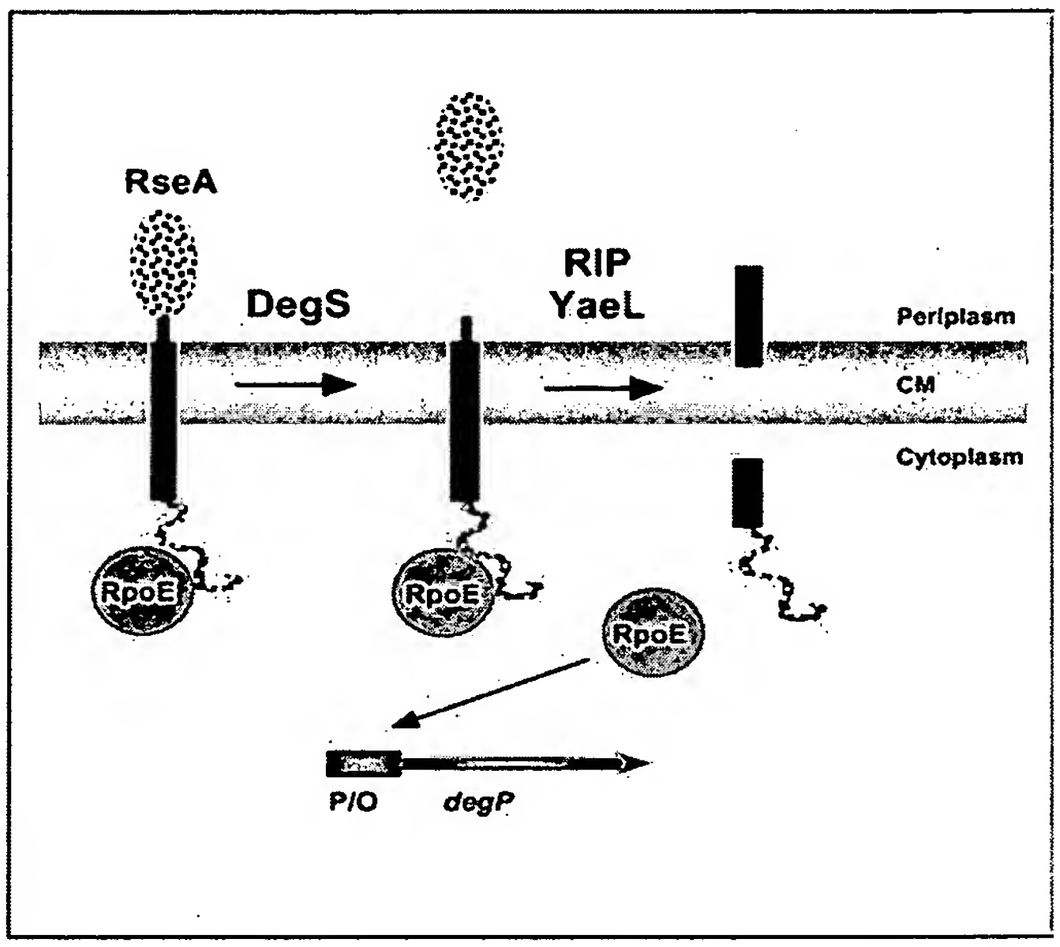
The antisigma activity of RseA is regulated on the level of protein stability. After an at least two-step proteolytic inactivation of RseA, σ^E is released to activate its target promoters (for review see 58). Two proteases, DegS and YaeL, were identified to be involved in the proteolytic inactivation of RseA (1, 38). The second step of RseA inactivation is carried out by YaeL. The metalloprotease YaeL is an integral cytoplasmic membrane protein belonging to the RIP family (37). YaeL-dependent proteolysis of RseA requires initial truncation by DegS, a member of the HtrA family of serine proteases (for review see 12). Elegant work by the Gross and Sauer groups showed that peptides corresponding to the C terminus of outer-membrane proteins activate the protease activity of DegS by binding to its PDZ domains (78). Activated DegS initiates cleavage of RseA and the σ^E cascade. Thus, when misfolded outer membrane proteins accumulate in the periplasm, they are thought to serve as signals for folding stress and thus activate the σ^E pathway. Promoters of the σ^E pathway include the *degP* and *fkpA* promoters, *rpoH* encoding the cytoplasmic heat shock sigma factor as well as *rpoE* and *rseABC* themselves (for review see 58). As there are two consecutive proteolytic events involved in RpoE activation, we discuss each protease individually.

The DegS Protease, a Prototype for Reversible Zymogen Activation

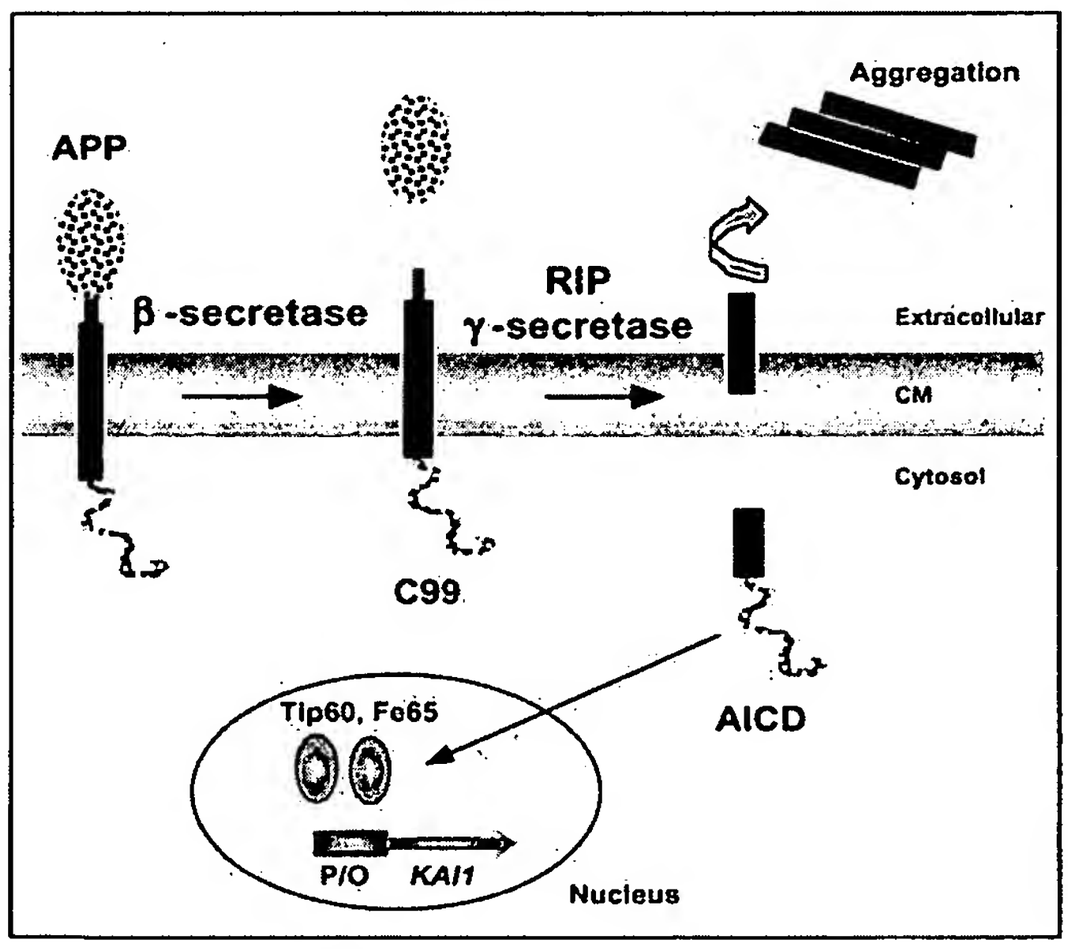
DegS, like all other members of the HtrA family, has a catalytic serine protease domain and a C-terminal PDZ domain (12). PDZ domains are protein modules that mediate specific protein-protein interactions and bind preferentially to the C-terminal 3–4 residues of the target protein (65). DegS belongs to the DegS/HtrA2 subfamily of HtrA proteases that share an N-terminal transmembrane segment, a trimeric architecture, and an involvement in regulatory processes. Human HtrA2 was recently implicated in apoptosis (12) and neuromuscular disorder (35), whereas DegS appears to be an extracytoplasmic stress sensor.

DegS differs from hexameric HtrAs such as DegP. DegP has very little substrate specificity, a strong preference for unfolded substrates, and switches from chaperone to protease in a temperature-dependent manner (69). In contrast, only one substrate, RseA, is known for DegS; RseA is cleaved once by DegS (1, 38, 78). HtrA proteases seem to reversibly switch between the inactive and active states, employing a mechanism that allows cells to rapidly respond to the need for proteolytic activity. In agreement with this model, the active sites of human HtrA2 and

A



B



E. coli DegP are present in the inactive state in their respective crystal structures (47, 67). Also, purified DegS is proteolytically inactive in the absence of activating peptides (78). To better understand how the switch in activity is effected, DegS has been crystallized in the presence and absence of activating peptides (81).

The protease domain of DegS is similar to other proteases of the trypsin family consisting of two perpendicular β-barrel lobes with a C-terminal helix. The catalytic triad (His96, Asp126, Ser201) is located in the crevice between the two β-barrels. Recent structural studies of HtrA proteins indicate that the protease domains form the rigid part of the structure, whereas the mobile PDZ domains are connected by a flexible linker peptide. Thus, the en-bloc mobility of the PDZ domains seems to represent an essential mechanistic feature. In DegS, however, the relative orientation of the PDZ and protease domains is fixed by an extended C terminus that promotes additional interactions between both domains. It appears that a precise positioning of the PDZ domain is crucial for the function of DegS. In contrast to previous models, the active sites of the DegS trimer are freely accessible and open up into the periplasmic space (Figure 2). Thus, the overall architecture of DegS differs considerably from that of other HtrA structures.

In the absence of bound activator, the proteolytic site of DegS is present in an inactive conformation, in which substrate binding as well as catalysis are prevented. The fine-tuned serine-histidine interplay of the catalytic triad seems to be severely disturbed, the oxyanion hole is misformed, the specificity pockets are occluded by several residues, and the structural feature required to bind the main-chain of substrates is absent.

The activating peptide corresponding to the C terminus of the outer-membrane protein OmpC interacts with the PDZ domain of DegS in a β-augmentation process. However, one of the peptide residues also interacts with the protease domain, triggering rearrangement of a particular active site loop. Ultimately, binding of the activating peptide leads to a precise series of conformational changes that proceed in a domino-like manner from the PDZ domain to the proteolytic site. As

Figure 1 The bacterial RseA and the human APP pathways. (*A*) The RpoE signal transduction pathway is involved in regulating transcription of stress promoters such as *degP*. Misfolded polypeptides are sensed by DegS, which is activated to cleave RseA in its periplasmic domain. Subsequent cleavage of RseA occurs within the membrane and is carried out by the RIP protease YaeL. After proteolytic inactivation of DegS, the sigma factor RpoE is released from the membrane to activate transcription of its target promoters. (*B*) The APP pathway is initiated by β-secretase, which removes the large extracellular domain. Subsequent processing within the membrane is carried out by the RIP protease γ-secretase, generating two fragments. The N-terminal fragment Aβ is released from the membrane into the medium where it can aggregate. The C-terminal cleavage product APP-CTF gamma (AICD) translocates into the nucleus where interaction with other proteins such as Fe65 and Tip60 promotes transcriptional activation of target promoters such as KAI1. CM is the cytoplasmic membrane.

a result, the remodeled proteolytic site is now present in its active conformation with a functional catalytic triad, oxyanion hole, and substrate specificity pockets (Figure 3). Therefore, in DegS, the PDZ domain is not a simple protein-binding domain. It takes on a novel regulatory function by being directly involved in intra- and intermolecular signaling. The PDZ domain represents a binding platform for an allosteric activator, thereby coupling the binding of an unfolded protein with activation of proteolytic activity.

Taken together, activation of DegS represents a novel mechanism with some similarities to the classical trypsin activation. Here, the transition from the inactive to an active state is achieved by reorientation of the so-called activation domain, which is also present in DegS. A further resemblance to trypsin activation is a disorder-order transition. In both cases, the activation domain is highly flexible in the inactive form, but becomes well defined in the active state. However, different activation signals are used. In trypsin, activation involves binding of a hydrophobic N terminus to the so-called Ile16-pocket and requires either propeptide cleavage or activators acting *in trans*. In DegS, peptide binding to the PDZ domain causes proteolytic activation. The remarkable difference is that activation of DegS is reversible, allowing cells to rapidly respond to the various folding stresses in a flexible manner.

The RIP Protease YaeL

Initial trimming of RseA by DegS involving removal of the periplasmic domain is required for subsequent processing by YaeL. Like DegS, YaeL is an essential protease (16, 37). It has four transmembrane segments and periplasmic N and C termini (20, 37). In addition to metallo protease consensus motifs, it also has a PDZ domain. In contrast to DegS, which has a C-terminal PDZ domain, the periplasmic PDZ domain of YaeL is located internally, i.e., between transmembrane segments 2 and 3 (37). YaeL belongs to the superfamily of PDZ-proteases. Its members include the *E. coli* proteases DegP, DegQ, and Tsp; human HtrA2; *Baccillus subtilis* SpoIVB, and several additional members linked to the tricorn family suggested by bioinformatic analysis (53). Although the relevant protease domains belong to different classes, they share a homologous PDZ domain that appears to be involved in regulating protease activity.

YaeL derivatives lacking the PDZ domain complement a chromosomal *yaeL* deletion and have increased proteolytic activity (39). These data indicate that the PDZ domain of YaeL is involved in regulation of protease activity. This observation has also been made for DegS, where the PDZ domain inhibits protease function (78). Even more surprising was the finding that PDZ-less YaeL degrades full-length RseA in the absence of DegS. This result indicates that the PDZ domain of YaeL also acts as a determinant of substrate specificity. In addition, mutational analyses identified two Gln-rich regions (residues 162–169 and 190–200) in the periplasmic C terminus of RseA that are important in the failure of wild-type YaeL to process full-length RseA in the absence of DegS (39).

Taken together, these data suggest that the RseA/σ^E signaling pathway is tightly controlled by the action of two proteases, DegS and YaeL. Tight control is achieved by the reversible activation of protease activity of DegS and by the exclusion of full-length RseA from proteolytic inactivation via YaeL. PDZ domains of DegS and YaeL are key players in this process because they control protease function and act as specificity determinants.

AMYLOID PRECURSOR PROTEIN PROCESSING PATHWAY

The human APP protein is synthesized with an N-terminal signal sequence. Mature APP has a large extracytoplasmic domain, one transmembrane segment, and a relatively short intracellular domain. APP is initially processed by either α- or β-secretases into a large ectodomain and a C-terminal fragment that contains the transmembrane segment. Subsequent processing of the β-secretase product, which is a 99-residue long C-terminal APP fragment, is mediated by γ-secretase, yielding N-terminal fragments of 40 or 42 residues and a C-terminal fragment of 50 residues (Figure 1*B*). These C-terminal fragments are thought to be involved in signaling, whereas the N-terminal fragments, termed Aβ, can aggregate extracellularly, and convincing evidence suggests that these aggregates are involved in Alzheimer's disease (AD) (for review see 71).

β-Secretase

The β-secretase (BACE1) cleaves APP between Met671/Asp672 located 28 residues upstream of the transmembrane segment of APP (for review see 34). BACE1 is a well-studied aspartic protease and its structure has been solved (30, 68). As classical aspartic protease inhibitors such as pepstatin do not inhibit β-secretase, β-secretase is believed to represent a novel class of membrane-bound aspartic proteases. Some APP mutants identified in patients suffering from early onset of AD, such as the Swedish variant, are cleaved at higher rates by β-secretase. In agreement with the model that β-secretase activity is critically involved in AD, BACE1 knockout mice produce no Aβ peptide (8, 64).

γ-Secretase

The γ-secretase is composed of at least 4 polytopic membrane proteins, presenilin (PS1), nicastrin, Aph1, and Pen-2 (21, 43, 72). Topological studies indicate that PS1, nicastrin, Aph1, and Pen-2 have 8, 1, 7, and 2 transmembrane segments, respectively (13, 19, 23, 84). Although strong genetic and biochemical evidence suggests that PS1 is the enzymatically active component of the complex, the precise function of the other components is unknown. PS1 does not contain a D(T/S)G(T/S) signature of classical aspartic proteases but has two highly conserved Asp residues in transmembrane segments 6 and 7 and a GxGD motif in transmembrane segment 7 that is also present in other RIP proteases (for review see 29).

A perhaps unique aspect of PS1 is that it can be processed into stable and nondissociating N- and C-terminal fragments (50, 55, 74). This proteolytic processing appears to be required for γ-secretase activity. Processing occurs in the large fourth cytoplasmic segment that connects transmembrane segments 6 and 7 containing the conserved Asp residues. The relevant part of the fourth cytoplasmic segment is encoded by exon 9. Exon 9 deletions possess γ-secretase activity in the absence of processing (28). Therefore, processing of PS1 could represent one way of regulating γ-secretase activity, although the mechanism is not understood and the relevant presenilase has not been conclusively identified. One attractive speculation might be that exon 9 contains an aspartic protease inhibitor that must be destroyed by processing to obtain active γ-secretase. If this model is correct, a zymogen-like mechanism of activation could apply for PS1.

A GENETIC SYSTEM TO MONITOR AND INVESTIGATE γ-SECRETASE FUNCTION AND ASSEMBLY A simple genetic reporter system for γ-secretase activity using yeast as a model organism has been developed (70). The system involves coexpression of γ-secretase and a C-terminal APP fragment of 55 residues, including the transmembrane segment, fused to the Gal4 transcriptional activator. Proteolytic processing of the reporter protein releases the Gal4 domain from the membrane from where it translocates to the nucleus and activates transcription of the promoter that drives expression of the reporter gene *lacZ* (Figure 4). *lacZ*, encoding β-galactosidase, provides a quantitative output signal that correlates with the activity of γ-secretase (21). This system or variations of it might allow mutational approaches to study such important questions as substrate specificity and the identification of processes that regulate γ-secretase activity. For example, random mutagenesis of genes encoding γ-secretase components could be used to identify which components are involved in excluding full-length APP from processing.

COMMON FEATURES OF THE RseA AND APP PATHWAYS

Numerous similarities can be observed between the human APP and the bacterial RseA pathway (Figure 1). Both APP and RseA are monotopic membrane proteins. DegS and β-secretase have one transmembrane segment, have extracytoplasmically localized active sites, and initiate proteolytic processing of RseA and APP, respectively. They are believed to be rate limiting as without their action no further processing of the substrate takes place. Another similarity is that YaeL and γ-secretase are RIP proteases, cleaving their substrate within the membrane. Finally, both substrates undergoing regulated intramembrane proteolysis are involved in signaling. RseA is part of the signal transduction cascade that monitors and responds to protein folding stress. The final product of β- and γ-secretases is a C-terminal APP fragment that is found in the nucleus (14, 25), where it interacts with a transcriptionally active complex containing the nuclear adaptor protein Fe65 and the histone deacetyltransferase TIP60 (10, 44). This complex targets, for

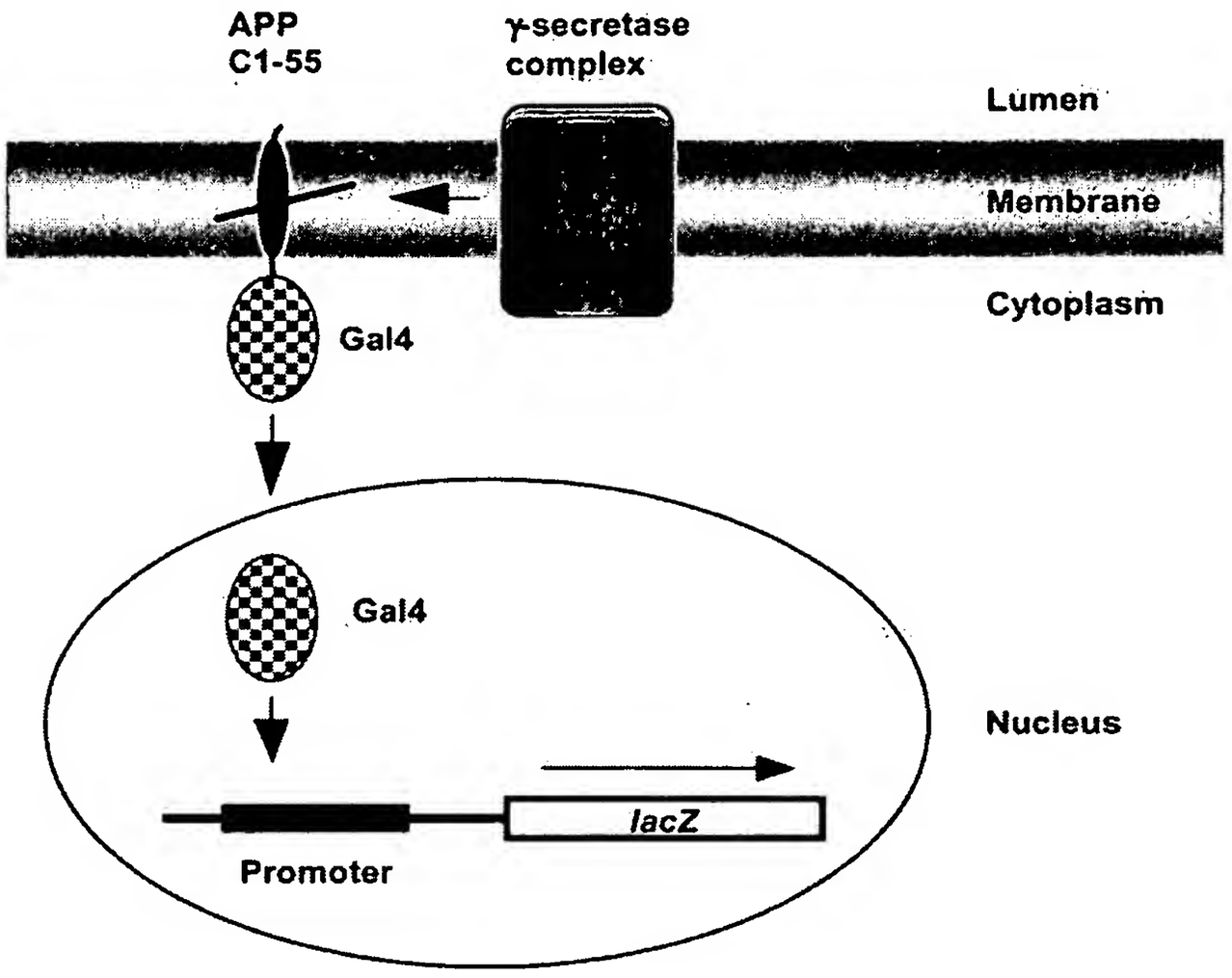


Figure 4 A genetic reporter system for γ-secretase activity. Expression of γ-secretase in yeast leads to intramembrane processing of an artificial model substrate composed of 55 residues of APP, including its transmembrane segment, fused to the Gal4 transcriptional activator domain. Proteolytic processing releases Gal4 from the membrane, followed by translocation to the nucleus where the transcription of the reporter gene *lacZ* is activated.

example, the *KAI1* promoter (3). As a cell surface molecule, KAI1 interacts with plasma membrane receptors and functions as a tumor metastasis suppressor (51). However, a recent study questions whether the C-terminal APP fragment has to be translocated to the nucleus or if it might interact and activate Fe65 while it is still associated with membrane (11), highlighting the need for more work to fully understand how APP fragments activate transcription.

INVOLVEMENT OF γ-SECRETASE AND PS1 IN SIGNALING

γ-secretase has been implicated in various other signaling cascades because C99 is not its only substrate. Other substrates include amyloid precursor-like proteins 1 and 2, Notch and its ligands Delta1 and Jagged2, Erb4, CD44, E- and N- cadherin, Nectin-1α, low-density lipoprotein-related protein (LRP), and syndecan 3 (for review see 22; 32, 42, 52, 66, 77). Although γ-secretase is involved in a variety

of signaling events that require RIP, there is also an interesting link between the APP and LRP pathways. Like APP, the LRP transmembrane receptor is cleaved by γ-secretase, and the generated C-terminal LRP fragment associates with Tip60 and Fe65. Therefore, this LRP fragment competes with the APP-derived/Fe65 transactivation mediated by Tip60 (45). In addition to its contribution to γ-secretase activity, PS1 is involved in additional regulatory processes. These include a PS1 function as a negative regulator of β-catenin by facilitating its phosphorylation and degradation (40).

CONCLUSIONS

The recruitment of proteases to signal transduction provides control of key elements on the level of stability. Studying proteolysis as a regulatory mechanism has revealed exciting and novel biology. A new class of proteases that are involved in numerous signaling cascades mediates regulated proteolysis that occurs within the membrane. Even though the individual proteases belong to different families, the underlying principles appear to be evolutionarily conserved. In addition, at least some proteases involved in regulation can switch their enzymatic activity on and off reversibly. This rapid adaptation of protease activity to the needs of signal processing uses PDZ domains as regulatory elements and as specificity determinants. These novel mechanisms have wide biological implications, and future studies will contribute to our understanding of how proteases regulate important biological processes.

ACKNOWLEDGMENTS

We thank Johannes Tkadletz for help with the figures. Work in the authors' laboratories is funded by BBSRC (ME) and by Boehringer Ingelheim (TC).

The *Annual Review of Genetics* is online at http://genet.annualreviews.org

LITERATURE CITED

1. Alba BM, Leeds JA, Onufryk C, Lu CZ, Gross CA. 2002. DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev.* 16:2156–68
2. Arsene F, Tomoyasu T, Bukau B. 2000. The heat shock response of *Escherichia coli*. *Int. J. Food Microbiol.* 55:3–9
3. Baek SH, Ohgi KA, Rose DW, Koo EH, Glass CK, Rosenfeld MG. 2002. Exchange of N-CoR corepressor and Tip60 coactivator complexes links gene expression by NF-κB and β-amyloid precursor protein. *Cell* 110:55–67
4. Bernstein HD. 2000. The biogenesis and assembly of bacterial membrane proteins. *Curr. Opin. Microbiol.* 3:203–9
5. Bode W, Huber R. 2000. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta* 1477:241–52

6. Brass JM, Higgins CF, Foley M, Rugman PA, Birmingham J, Garland PB. 1986. Lateral diffusion of proteins in the periplasm of *Escherichia coli*. *J. Bacteriol.* 165:787–95
7. Brown MS, Ye J, Rawson RB, Goldstein JL. 2000. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* 100:391–98
8. Cai H, Wang Y, McCarthy D, Wen H, Borchelt DR, et al. 2001. BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat. Neurosci.* 4:233–34
9. Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, et al. 2003. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell* 11:1067–78
10. Cao X, Sudhof TC. 2001. A transcriptionally (correction of transcriptively) active complex of APP with Fe65 and histone acetyltransferase Tip60. *Science* 293:115–20
11. Cao X, Sudhof TC. 2004. Dissection of APP-dependent transcriptional transactivation. *J. Biol. Chem.* 279:24601–11
12. Clausen T, Southan C, Ehrmann M. 2002. The HtrA family of proteases. Implications for protein composition and cell fate. *Mol. Cell* 10:443–55
13. Crystal AS, Morais VA, Pierson TC, Pijak DS, Carlin D, et al. 2003. Membrane topology of gamma-secretase component PEN-2. *J. Biol. Chem.* 278:20117–23
14. Cupers P, Orlans I, Craessaerts K, Annaert W, De Strooper B. 2001. The amyloid precursor protein (APP)-cytoplasmic fragment generated by gamma-secretase is rapidly degraded but distributes partially in a nuclear fraction of neurones in culture. *J. Neurochem.* 78:1168–78
15. Danese PN, Silhavy TJ. 1998. Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annu. Rev. Genet.* 32:59–94
16. Dartigalongue C, Missiakas D, Raina S. 2001. Characterization of the *Escherichia coli* sigmaE regulon. *J. Biol. Chem.* 276: 20866–75
17. De Las Penas A, Connolly L, Gross CA. 1997. SigmaE is an essential sigma factor in *Escherichia coli*. *J. Bacteriol.* 179: 6862–64
18. Degterev A, Boyce M, Yuan J. 2003. A decade of caspases. *Oncogene* 22:8543–67
19. Doan A, Thinakaran G, Borchelt DR, Slunt HH, Ratovitsky T, et al. 1996. Protein topology of presenilin 1. *Neuron* 17: 1023–30
20. Drew D, Sjostrand D, Nilsson J, Urbig T, Chin CN, et al. 2002. Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc. Natl. Acad. Sci. USA* 99:2690–95
21. Edbauer D, Winkler E, Regula JT, Pesold B, Steiner H, Haass C. 2003. Reconstitution of gamma-secretase activity. *Nat. Cell Biol.* 5:486–88
22. Fortini ME. 2002. Gamma-secretase-mediated proteolysis in cell-surface-receptor signalling. *Nat. Rev. Mol. Cell Biol.* 3:673–84
23. Fortna RR, Crystal AS, Morais VA, Pijak DS, Lee VM, Doms RW. 2004. Membrane topology and nicastrin-enhanced endoproteolysis of APH-1, a component of the gamma-secretase complex. *J. Biol. Chem.* 279:3685–93
24. Friedrich R, Panizzi P, Fuentes-Prior P, Richter K, Verhamme I, et al. 2003. Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation. *Nature* 425:535–39
25. Gao Y, Pimplikar SW. 2001. The gamma-secretase-cleaved C-terminal fragment of amyloid precursor protein mediates signaling to the nucleus. *Proc. Natl. Acad. Sci. USA* 98:14979–84
26. Gottesman S. 1987. Regulation by proteolysis. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular*

Biology, ed. C Neidhart, J Ingraham, KB Low, B Magasanik, M Schaechter, HE Umbarger, pp. 1308–12. Washington, DC: ASM
27. Gottesman S. 2003. Proteolysis in bacterial regulatory circuits. *Annu. Rev. Cell Dev. Biol.* 19:565–87
28. Haass C, De Strooper B. 1999. The presenilins in Alzheimer's disease-proteolysis holds the key. *Science* 286:916–19
29. Haass C, Steiner H. 2002. Alzheimer disease gamma-secretase: a complex story of GxGD-type presenilin proteases. *Trends Cell Biol.* 12:556–62
30. Hong L, Koelsch G, Lin X, Wu S, Terzyan S, et al. 2000. Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. *Science* 290:150–53
31. Huang X, Knoell CT, Frey G, Hazegh-Azam M, Tashjian AH Jr, et al. 2001. Modulation of recombinant human prostate-specific antigen: activation by Hofmeister salts and inhibition by azapeptides. *Biochemistry* 40:11734–41
32. Ikeuchi T, Sisodia SS. 2003. The Notch ligands, Delta1 and Jagged2, are substrates for presenilin-dependent "gamma-secretase" cleavage. *J. Biol. Chem.* 278: 7751–54
33. Jenal U, Hengge-Aronis R. 2003. Regulation by proteolysis in bacterial cells. *Curr. Opin. Microbiol.* 6:163–72
34. John V, Beck JP, Bienkowski MJ, Sinha S, Heinrikson RL. 2003. Human beta-secretase (BACE) and BACE inhibitors. *J. Med. Chem.* 46:4625–30
35. Jones JM, Datta P, Srinivasula SM, Ji W, Gupta S, et al. 2003. Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice. *Nature* 425:721–27
36. Kadokura H, Katzen F, Beckwith J. 2003. Protein disulfide bond formation in prokaryotes. *Annu. Rev. Biochem.* 72: 111–35
37. Kanehara K, Akiyama Y, Ito K. 2001. Characterization of the *yaeL* gene product and its S2P-protease motifs in *Escherichia coli*. *Gene* 281:71–79
38. Kanehara K, Ito K, Akiyama Y. 2002. YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev.* 16:2147–55
39. Kanehara K, Ito K, Akiyama Y. 2003. YaeL proteolysis of RseA is controlled by the PDZ domain of YaeL and a Gln-rich region of RseA. *EMBO J.* 22:6389–98
40. Kang DE, Soriano S, Xia X, Eberhart CG, De Strooper B, et al. 2002. Presenilin couples the paired phosphorylation of beta-catenin independent of axin: implications for beta-catenin activation in tumorigenesis. *Cell* 110:751–62
41. Khan AR, James MN. 1998. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* 7:815–36
42. Kim DY, Ingano LA, Kovacs DM. 2002. Nectin-1alpha, an immunoglobulin-like receptor involved in the formation of synapses, is a substrate for presenilin/gamma-secretase-like cleavage. *J. Biol. Chem.* 277:49976–81
43. Kimberly WT, LaVoie MJ, Ostaszewski BL, Ye W, Wolfe MS, Selkoe DJ. 2003. Gamma-secretase is a membrane protein complex comprised of presenilin, nicastrin, Aph-1, and Pen-2. *Proc. Natl. Acad. Sci. USA* 100:6382–87
44. Kimberly WT, Zheng JB, Guenette SY, Selkoe DJ. 2001. The intracellular domain of the beta-amyloid precursor protein is stabilized by Fe65 and translocates to the nucleus in a notch-like manner. *J. Biol. Chem.* 276:40288–92
45. Kinoshita A, Shah T, Tangredi MM, Strickland DK, Hyman BT. 2003. The intracellular domain of the low density lipoprotein receptor-related protein modulates transactivation mediated by amyloid precursor protein and Fe65. *J. Biol. Chem.* 278:41182–88
46. Koebnik R, Locher KP, Van Gelder P. 2000. Structure and function of bacterial

outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37:239–53
47. Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T. 2002. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* 416:455–59
48. Larsen CN, Finley D. 1997. Protein translocation channels in the proteasome and other proteases. *Cell* 91:431–34
49. Lazure C. 2002. The peptidase zymogen proregions: nature's way of preventing undesired activation and proteolysis. *Curr. Pharm. Des.* 8:511–31
50. Levitan D, Lee J, Song L, Manning R, Wong G, et al. 2001. PS1 N- and C-terminal fragments form a complex that functions in APP processing and Notch signaling. *Proc. Natl. Acad. Sci. USA* 98:12186–90
51. Maecker HT, Todd SC, Levy S. 1997. The tetraspanin superfamily: molecular facilitators. *FASEB J.* 11:428–42
52. Marambaud P, Wen PH, Dutt A, Shioi J, Takashima A, et al. 2003. A CBP binding transcriptional repressor produced by the PS1/epsilon-cleavage of N-cadherin is inhibited by PS1 FAD mutations. *Cell* 114:635–45
53. Pallen MJ, Lam AC, Loman N. 2001. Tricorn-like proteases in bacteria. *Trends Microbiol.* 9:518–21
54. Patil C, Walter P. 2001. Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.* 13:349–55
55. Podlisny MB, Citron M, Amarante P, Sherrington R, Xia W, et al. 1997. Presenilin proteins undergo heterogeneous endoproteolysis between Thr291 and Ala299 and occur as stable N- and C-terminal fragments in normal and Alzheimer brain tissue. *Neurobiol. Dis.* 3: 325–37
56. Raffa RG, Raivio TL. 2002. A third envelope stress signal transduction pathway in *Escherichia coli*. *Mol. Microbiol.* 45:1599–611
57. Raivio TL, Silhavy TJ. 1999. The sigmaE and Cpx regulatory pathways: overlapping but distinct envelope stress responses. *Curr. Opin. Microbiol.* 2:159–65
58. Raivio TL, Silhavy TJ. 2001. Periplasmic stress and ECF sigma factors. *Annu. Rev. Microbiol.* 55:591–624
59. Rawlings N, Barrett A. 1994. Families of serine peptidases. *Methods Enzymol.* 244:19–61
60. Rawlings ND, Barrett AJ. 1994. Families of cysteine peptidases. *Methods Enzymol.* 244:461–86
61. Rawlings ND, Barrett AJ. 1995. Evolutionary families of metallopeptidases. *Methods Enzymol.* 248:183–228
62. Rawlings ND, Barrett AJ. 1995. Families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods Enzymol.* 248:105–20
63. Rawlings ND, Tolle DP, Barrett AJ. 2004. MEROPS: the peptidase database. *Nucleic Acids Res.* 32 Database issue: D160-4
64. Roberds SL, Anderson J, Basi G, Bienkowski MJ, Branstetter DG, et al. 2001. BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. *Hum. Mol. Genet.* 10: 1317–24
65. Saras J, Heldin CH. 1996. PDZ domains bind carboxy-terminal sequences of target proteins. *Trends Biochem. Sci.* 21:455–58
66. Schulz JG, Annaert W, Vandekerckhove J, Zimmermann P, De Strooper B, David G. 2003. Syndecan 3 intramembrane proteolysis is presenilin/gamma-secretase-dependent and modulates cytosolic signaling. *J. Biol. Chem.* 278:48651–57
67. Shi Y. 2002. Mechanisms of caspase activation and inhibition during apoptosis. *Mol. Cell* 9:459–70
68. Sinha S, Anderson JP, Barbour R, Basi GS, Caccavello R, et al. 1999. Purification

and cloning of amyloid precursor protein beta-secretase from human brain. *Nature* 402:537–40
69. Spiess C, Beil A, Ehrmann M. 1999. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97:339–47
70. Steiner H, Pesold B, Haass C. 1999. An in vivo assay for the identification of target proteases which cleave membrane-associated substrates. *FEBS Lett.* 463:245–49
71. Suh Y, Checler F. 2002. Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. *Pharmacol. Rev.* 54: 469–525
72. Takasugi N, Tomita T, Hayashi I, Tsuruoka M, Niimura M, et al. 2003. The role of presenilin cofactors in the gamma-secretase complex. *Nature* 422:438–41
73. Tamm LK, Arora A, Kleinschmidt JH. 2001. Structure and assembly of beta-barrel membrane proteins. *J. Biol. Chem.* 276:32399–402
74. Thinakaran G, Borchelt DR, Lee MK, Slunt HH, Spitzer L, et al. 1996. Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. *Neuron* 17:181–90
75. Urban S, Freeman M. 2002. Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr. Opin. Genet. Dev.* 12:512–18
76. Urban S, Lee JR, Freeman M. 2001. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* 107:173–82
77. Walsh DM, Fadeeva JV, LaVoie MJ, Paliga K, Eggert S, et al. 2003. gamma-Secretase cleavage and binding to FE65 regulate the nuclear translocation of the intracellular C-terminal domain (ICD) of the APP family of proteins. *Biochemistry* 42:6664–73
78. Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT. 2003. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* 113:61–71
79. Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B. 2002. Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* 296:2215–18
80. Weihofen A, Martoglio B. 2003. Intramembrane-cleaving proteases: controlled liberation of proteins and bioactive peptides. *Trends Cell Biol.* 13:71–78
81. Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T. 2004. Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease domain. *Cell* 17:483–94
82. Wülfing C, Plückthun A. 1994. Protein folding in the periplasm of *E. coli*. *Mol. Microbiol.* 12:685–92
83. Ye S, Goldsmith EJ. 2001. Serpins and other covalent protease inhibitors. *Curr. Opin. Struct. Biol.* 11:740–45
84. Yu G, Nishimura M, Arawaka S, Levitan D, Zhang L, et al. 2000. Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature* 407:48–54

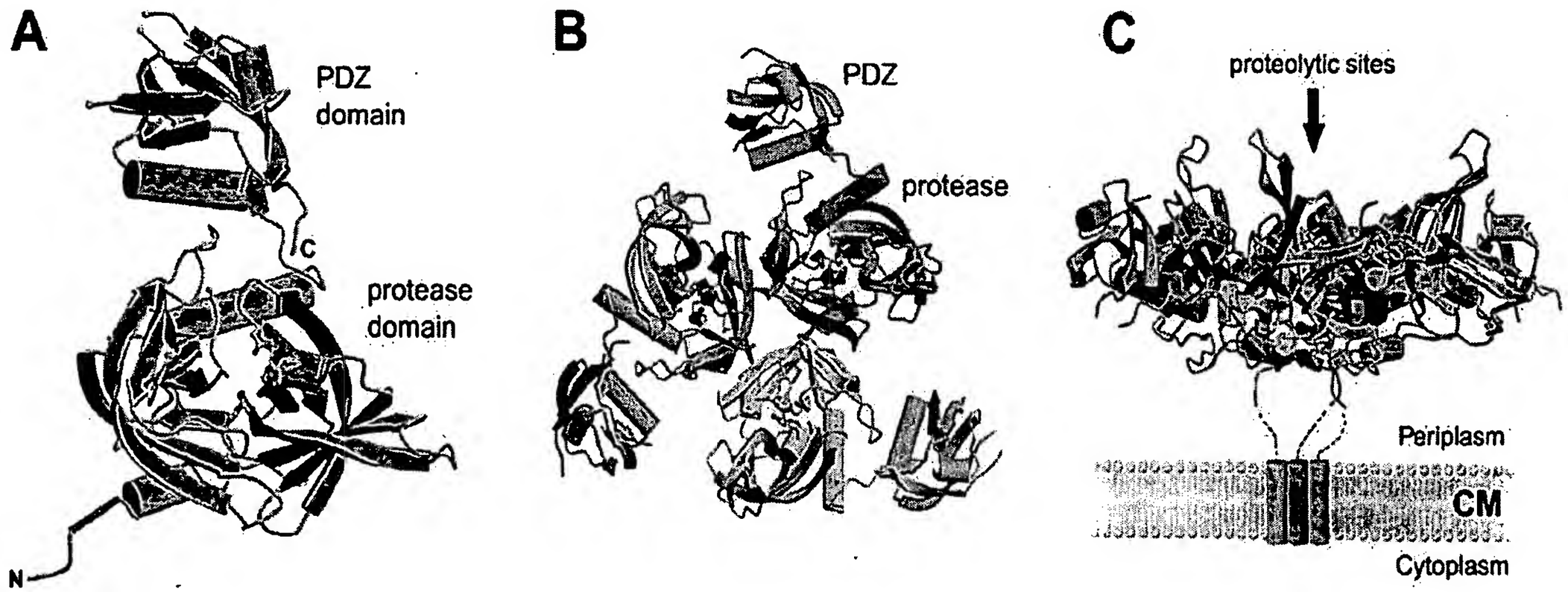


Figure 2 Structure of *E. coli* DegS. (*A*) Ribbon presentation of the monomer with labeled N terminus and C terminus. The catalytic triad is shown as a stick model. (*B*) Top view on the DegS trimer with each subunit colored differently. (*C*) Side view of the trimer illustrating the relative orientation of DegS in the periplasm.

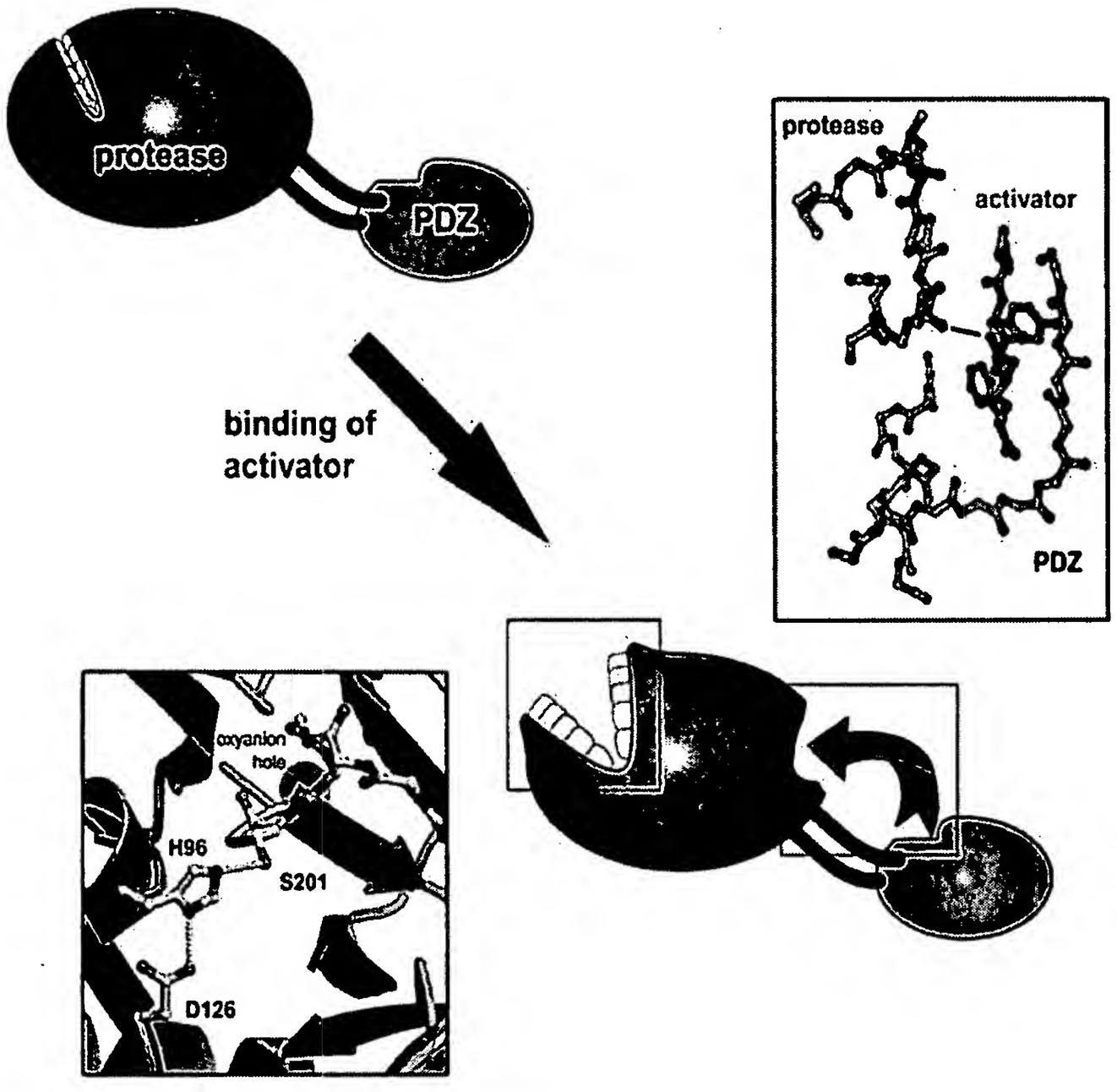


Figure 3 Scheme of DegS activation. In DegS, the PDZ domain obtains a regulatory function, offering a binding site for an allosteric activator. One of the highlighted regions illustrates binding of the activator to PDZ and protease domain. While the backbone of the activator is anchored to the PDZ domain, one of its side chains undergoes a specific interaction with a protease active site loop. This interaction triggers rearrangement of the activation domain and formation of a functional catalytic site, as shown in the second illustration.

Annual Review of Genetics
Volume 38, 2004

CONTENTS

PROTEASE DEGRADOMICS: A NEW CHALLENGE FOR PROTEOMICS

Carlos López-Otín and Christopher M. Overall‡*

Degradomics — the application of genomic and proteomic approaches to identify the protease and protease-substrate repertoires, or 'degradomes', on an organism-wide scale — promises to uncover new roles for proteases *in vivo*. This knowledge will facilitate the identification of new pharmaceutical targets to treat disease. Here, we review emerging degradomic techniques and concepts.

PROTEASE
An enzyme that cleaves proteins by the catalysis of peptide-bond hydrolysis. On the basis of their catalytic mechanism, proteases belong to one of five classes (aspartic, cysteine, metallo, serine, or threonine).

PROTEOLYTIC PROCESSING
Proteolysis that is distinct from degradation in that it represents highly specific and limited substrate cleavage, which results in a specific change of protein function.

**Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain.*
‡Departments of Biochemistry and Molecular Biology, and Oral Biological and Medical Sciences, Protein Engineering Network of Centers of Excellence, University of British Columbia, Vancouver British Columbia, V6T 1Z3 Canada. Correspondence to C.M.O.
e-mail: chris.overall@ubc.ca
doi:10.1038/nrm858

Our view of the proteolytic universe has expanded considerably in recent years. PROTEASES were initially characterized as nonspecific degradative enzymes that are associated with protein catabolism. However, it is becoming increasingly recognized that proteolysis represents another mechanism for achieving precise cellular control of biological processes in all living organisms, through the highly specific hydrolysis of peptide bonds[1]. This highly specific and limited substrate cleavage is termed PROTEOLYTIC PROCESSING. Proteases, through their ability to catalyse irreversible hydrolytic reactions, regulate the fate and activity of many proteins by controlling appropriate intra- or extracellular localization; shedding from cell surfaces; activation or inactivation of proteases and other enzymes, cytokines, hormones or growth factors; conversion of receptor agonists to antagonists; and exposure of cryptic neoproteins (which is when the proteolytic cleavage products are functional proteins with roles that are distinct from the parent molecule). Hence, proteases initiate, modulate and terminate a wide range of important cellular functions by processing bioactive molecules, and thereby directly control essential biological processes, such as DNA replication, cell-cycle progression, cell proliferation, differentiation and migration, morphogenesis and tissue remodelling, neuronal outgrowth, haemostasis, wound healing, immunity, angiogenesis and apoptosis[1,2].

Considering the functional relevance of proteases for all living processes, including cell death, it is not difficult to understand that a deficiency, or a misdirected temporal and spatial activity, of these enzymes underlies several pathological conditions such as cancer, arthritis, neurodegenerative and cardiovascular diseases[1,2]. Moreover, many infectious microorganisms, viruses and parasites use proteases as virulence factors, and animal venom commonly contains proteases to effect tissue destruction or to evade host responses. Accordingly, many proteases or their substrates are an important focus of attention for the pharmaceutical industry as potential drug targets.

Owing to the expanding roles for proteolytic enzymes, there has been an increasing interest in the identification and functional characterization of the many proteases that are present in various organisms, from bacteria to man. The near completion of several large-scale genome-sequencing projects has provided new opportunities to appreciate the complexity of protease systems. According to our survey of the human genome, more than 500 genes that encode proteases or protease-like molecules comprise our proteolytic labyrinth. The most recent release (17 December 2001) of the protease database — MEROPS version 5.7 — lists 461 proteases and homologues in man, with the metalloproteinases and serine proteases comprising the largest classes, having at present 159 and 140 members, respectively (TABLE 1). As unidentified genes are characterized and 'hidden' proteases, such as Rhomboid[3], are recognized — some of which might have new catalytic machinery and highly specific activities — the total number of proteases in humans could grow beyond 500. The task of characterizing both the functions of all proteases (existing and newly identified), and their functional connections with other proteases and inhibitors in the various protease systems in humans, is therefore daunting, and model organisms will be invaluable for this. The murine protease map is similar to the human

MATRIX METALLOPROTEINASES
A family of 23 endoproteinases in humans that are encoded by 24 genes. These are characterized by a HEXXHXXGXXH zinc-binding motif, a cysteine-switch mechanism of proenzyme latency, an ability to cleave extracellular-matrix and bioactive molecules, and inhibition by tissue inhibitors of metalloproteinases (TIMPs).

Table 1 | **Current numbers of proteases in humans and model species**

		Catalytic class of protease				
Species	**Total***	**Aspartic**	**Cysteine**	**Metallo**	**Serine**	**Threonine**
Homo sapiens	461	18	121	159	140	23
Caenorhabditis elegans	353	26	93	151	62	21
Drosophila melanogaster	513	38	59	157	225	34
Mus musculus	383	11	93	120	137	22
Rattus norvegicus	227	10	41	77	80	19

*The total numbers are accurate as of the date of release of the most recent MEROPS database v 5.7 (17 December 2001), but will continue to grow as new proteases are discovered and characterized in the existing databases and in new sequence deposits.

map, although there are families such as the kallikreins, cathepsins or MATRIX METALLOPROTEINASES (MMPs) that have evolved somewhat differently[4–6] (TABLE 1). Surprisingly, *Drosophila melanogaster*, which has a gene content that is considerably lower than these vertebrate organisms, shows a similar number of protease genes as a result of the impressive expansion of a group of trypsin-like serine proteases in the fly genome[7]. However, further studies are needed to clarify the genetic and molecular bases of the evolutionary differences between the protease repertoires of these organisms. So, in addition to the universal proteolytic 'routines' that are common to all organisms, there are also specific functions that are carried out by unique proteases in different species. Finally, and in keeping with the observed protease complexity, recent studies are uncovering many diverse new substrates and new endogenous inhibitors that have the ability to balance protease activity in physiological and pathological conditions. Hence, the emerging pattern in the proteolytic world is one of diversity and complexity, which produces new layers of control — often of an exquisite nature — for many pivotal cell and tissue functions through precise and limited proteolysis rather than through nonspecific, general protein catabolism.

Despite these advances, the substrates and *in vivo* roles for newly identified proteases are unknown and, even for proteases that have been well characterized, their biological functions are often not fully understood. New techniques are urgently required to identify the protease repertoire that is expressed and active in a cell, tissue or organism, as well as to identify all the natural substrates of each protease. This article addresses these crucial issues — the 'omics' language is adapted for proteases and the new terms are defined. We cover potential uses for global analyses of protease systems, which — as well as the limitations of existing strategies — will drive the development of new approaches and technologies for the system-wide analysis of proteases and their substrates. In particular, four classes of protease chips will be discussed, which can identify complete sets of proteases and their substrate repertoires at an organism, tissue or cellular level in a high-throughput manner. Other emerging techniques — such as the use of fluorogenic-labelled protease inhibitors, chemical proteomics and protease domains to identify new substrates — will also be presented to encourage the further development of these techniques. Finally, we discuss the impact that this new system-wide approach will have on the protease field in terms of our views of the roles of proteases *in vivo* in health and disease states, and in terms of the identification of new targets for the pharmaceutical industry.

Box 1 | **Genomics and proteomics: the parents of degradomics**

Degradome and degradomics are two terms that have been adapted from the fields of genomics and proteomics, and these terms are defined below together with other 'omics' terms that appear in this Review.

- *Genome.* The entire collection of genes in the complete DNA sequence of an organism.
- *Transcriptome.* The complete set of mRNAs that are transcribed from the genome.
- *Proteome.* The expressed set of proteins that are encoded by the genome.
- *Genomics.* Investigations and techniques for identifying the genome.
- *Proteomics.* Investigations and techniques for identifying the proteome.
- *Degradomics.* All genomic and proteomic investigations and techniques regarding the genetic, structural and functional identification and characterization of proteases, and their substrates and inhibitors, that are present in an organism.
- *Degradomes.* The complete set of proteases that are expressed at a specific time by a cell, tissue or organism. The degradome of a protease is its substrate repertoire (FIG. 1).

Degradomics and degradomes

Evidence for the increasing complexity and importance of the proteolytic systems that function in all organisms, and the ability to analyse systems in their entirety on genome- and proteome-wide scales (BOX 1), necessitates the introduction of new terms to clarify emerging concepts in this field. So, 'degradomics' was first coined to define the substrate repertoire of a protease on a proteome-wide scale[8]. Given the historical connection between proteases and general protein degradation, we propose that this is an appropriate and intuitive term, despite the more important, and now increasingly recognized, roles of proteases as processing enzymes. We extend this definition to one that describes all genomic and proteomic approaches for the identification and characterization of proteases that are present in an organism, including the substrates that are targeted by these proteases and their endogenous inhibitors (FIG. 1). We also propose the term 'degradome' to encompass two concepts. First, it is the complete set of proteases that are expressed at a specific moment or circumstance by a cell, tissue or organism. Second, the degradome of a protease is the complete natural substrate repertoire of that enzyme in a cell, tissue or organism (FIG. 1). The field of degradomics will be built using emerging, and new,

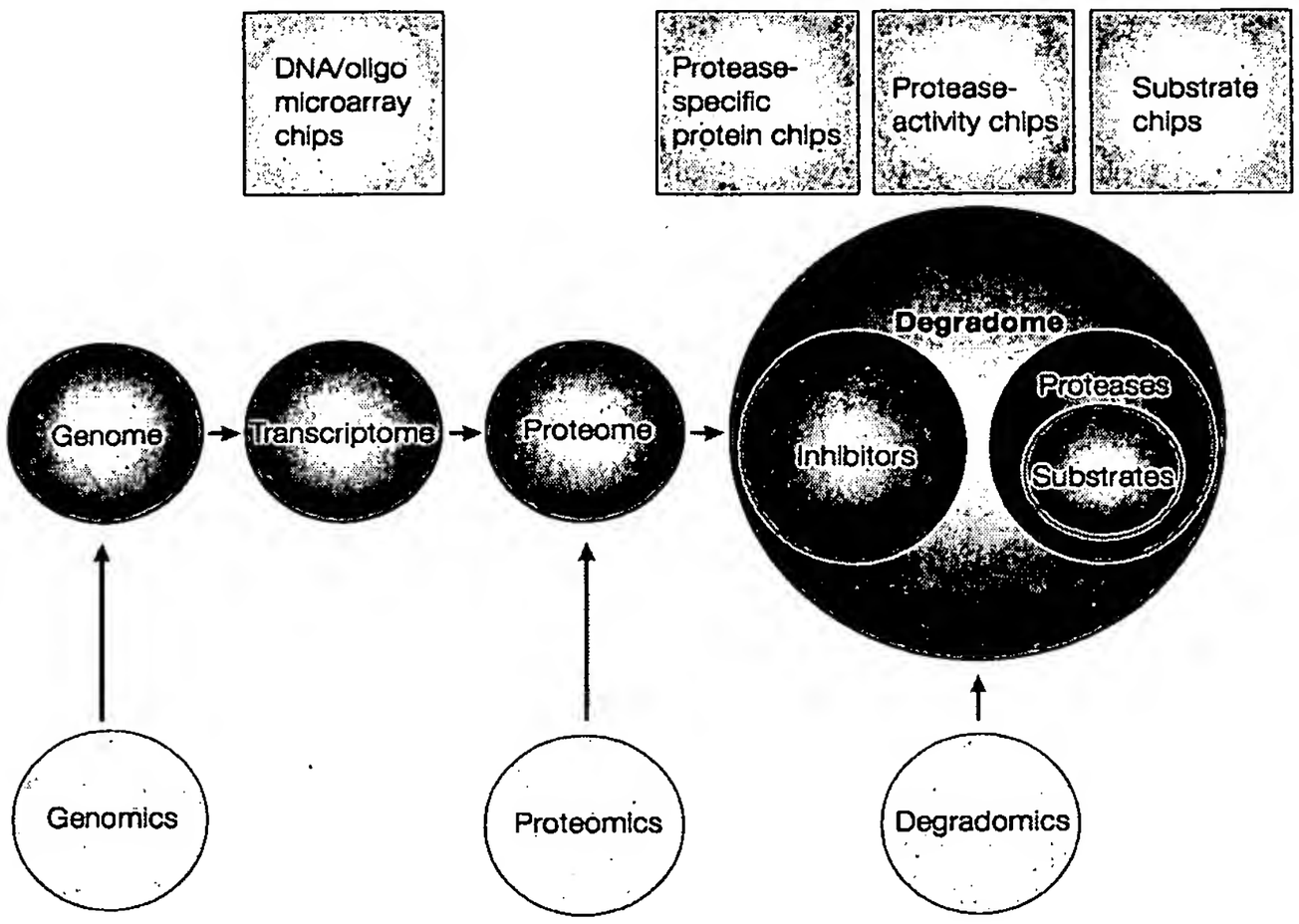


Figure 1 | **Relationship of degradomics to the fields of proteomics and genomics, and of the degradome to the proteome and genome.** Degradomics is the application of genomic and proteomic approaches to identify the protease and protease-substrate repertoires — or 'degradomes' — on an organism-wide scale. A glossary of definitions is provided in BOX 1. Examples of protease chips that can be used for degradomics investigations are also shown in the figure (shaded green).

genomic and proteomic technologies to investigate and define both types of protease degradome.

Identifying the substrate degradomes of individual proteases will facilitate our understanding of their physiological and pathological roles and thereby point to new drug targets. This information, in conjunction with knowledge of the protease degradome of a cell, will increase our understanding of the biological roles of proteases in the cellular context with respect to cell function and pathology. Similar information on a tissue-wide scale should prove useful in the molecular diagnosis of disease, with the calibration of protease levels to disease severity or tumour grade enabling more accurate prognostic predictions to be made for patients. The protease degradome of an organism will be largely defined by genomic and bioinformatic analyses of gene sequences, with the expression pattern of protease messenger RNAs — the transcriptome (BOX 1) — showing tissue- or cell-specific expression levels that reflect proteolytic potential. However, functional degradomics is required to determine the actual proteolytic activity that is expressed at a particular time by a cell or tissue. Functional degradomics has two branches: the first is based on activity profiling of individual proteases, and the second involves determination of the net cleavage of target substrates. So, instead of defining individual contributions by specific proteases, this latter aim considers the protease degradome as a system that leads to substrate cleavage. The field of degradomics promises to uncover new proteases and physiological substrates, and to identify new and known regulatory pathways that are controlled by proteolytic processing. The regulation of these pathways might be disrupted in disease states, or host proteases might be used by microorganisms for infection, and could therefore be therapeutically targeted.

MASS SPECTROMETRY
A technique that precisely measures sample mass from the analysis of mass-to-charge ratio (m/z).

From global concepts to global approaches

Proteases do not operate in isolation — they are expressed and function in the context of a proteolytic system that is comprised of related and unrelated proteases, their substrates and cleavage products, inhibitors, cell receptors and binding proteins. It is often the case that one protease can cleave many substrates, and also that many proteases can cleave the same substrate, activate other members of a proteolytic cascade, undergo autolysis *in cis* and *in trans*, and degrade other members of the system. Only by considering individual proteases as a part of a system, and the proteolytic systems as a whole, can the impact of the protease degradome on the substrate degradome *in vivo* be understood and hence its perturbations recognized in disease. As systems hold considerably more information than their individual components, the analysis of proteases needs to move from characterization of individual proteins to the system-wide level. However, in the rapidly approaching post-proteomic era, the detailed biological characterization of individual proteases and drug targets will still be important in order to exploit the wealth of information that is generated by degradomics analyses.

The degradomics challenge is therefore significant — even for well-studied proteases, their biological roles and their relationships with other components of the protease systems are at present not fully understood. Specific activity and redundancy are crucial determinants of the role of a particular protease in a proteolytic system. The hierarchical importance of proteases within a system is also affected by expression levels, temporal/spatial distribution, activation, turnover and inhibition — properties that profoundly influence proteolytic potential *in vivo*. This is a crucial issue for drug development, as several proteases might cleave the same substrate *in vitro*, but, *in vivo*, substrate cleavage might be restricted to one, or only a few, of these proteases.

Although new protease substrates are being recognized, they are typically discovered by serial approaches that are time consuming, usually cumbersome, and have incomplete coverage. Moreover, substrate identification *in vitro* does not necessarily indicate that the protein will be a biologically relevant substrate *in vivo*. Considering the number and diversity of proteases that are present in different organisms, innovative approaches and tools are needed to profile the expressed protease repertoire and to screen for new substrates. Distilling this information will require many tools and iterative approaches.

Conventional methods of proteomics at present involve denaturation of all proteins in a mixture by two-dimensional electrophoretic methods or by tryptic digestion and serial liquid-chromatography (LC) steps before MASS SPECTROMETRY, but considerable information is lost as a result of denaturation or digestion. Fortunately, innovative and rapid advances in proteomics for native protein analysis have been made by new or improved sample preparation, labelling and

Table 2 | **Overview of degradomics approaches**

Degradomics approach	Analysis scale	Target	Limitations
DNA microarray chips	Transcriptome	Messenger RNA	Expression levels do not relate to protein abundance. Does not reflect recruitment of proteases *in trans* from other cells or tissues.
Protease-specific protein chips	Proteome	Protease protein	Abundance does not necessarily reflect activity. Lack of specific probes for all proteases.
Protease-activity chips	Proteome	Active protease	Activity does not necessarily indicate substrate cleavage. Measures absolute levels without considering protease turnover.
Substrate chips	Proteome	Protease substrate	Does not identify the active proteases. Difficult to obtain proteome-wide protein arrays. Substrate protein might not be in the correct three-dimensional biological conformation.
Two-dimensional gel tandem mass spectrometry	Proteome	Protease substrate Protease	Low-mass cleavage fragments (<~8 kDa) and fragments with high or low isoelectric points are not resolved on two-dimensional gels, and thereby exclude many bioactive mediators. Membrane proteins, such as receptors and adhesion molecules, are difficult to study. Low abundance proteins are not detected because of sensitivity limits. Protein masses are only inaccurately determined by electrophoresis before in-gel tryptic digestion. Substrates and large cleavage products might not resolve well.
Inhibitor-based electrophoretic profiling	Proteome	Active protease	Lack of specific probes for all proteases. Limited throughput. Denaturing method, so not suitable for proteases without a covalent acyl intermediate.
Inhibitor- or antibody-based liquid chromatography and tandem mass spectrometry	Proteome	Active protease	Non-quantitative.
Chemical proteomics	Proteome	Active protease	Pharmacokinetic problems might limit effective concentration *in vivo*. Toxicity concerns. Small amounts of active protease might not be inhibited. Due to the difficulty in developing protease-specific inhibitors, related proteases might also be inhibited.
Targeted isotope-coded affinity tags	Proteome	Active or total protease depending on tag selection	Quantitative, but difficulties arise in developing protease-specific probes. Development of covalent probes is a huge hurdle for proteases without a substrate-cleavage acyl intermediate. The mass of the intact protein is not measured by mass spectrometry.

fractionation techniques before mass spectrometric analysis. Among the emerging technologies, which can also be applied to the global analysis of proteases, are protease chips, which we envisage can be divided into four different formats (TABLE 2) — conventional DNA microarray chips, protease-specific protein chips, protease-activity chips and substrate chips. The global approaches that are available to degradomics will now be discussed in more detail.

Protease DNA microarray chips. Conventional DNA microarrays are based on complementary DNA or oligonucleotide-specific probes for all the different proteases of that species, as defined by whole-genome sequencing projects and bioinformatic analyses. Ideally, these chips would also contain probes that are specific for the different protease inhibitors that are produced by an organism. These chips can be used to obtain a general view of the protease and inhibitor transcriptomes in a normal or pathological tissue and, with developments in technology, will eventually be used to obtain a specific view at the single-cell level. However, mRNA expression levels of protease genes do not accurately reflect the protein expression levels of these enzymes, nor do they show proteases that are recruited from remote sources, such as the serum or the adjacent tissue and infiltrating cells. DNA-based chips should therefore be complemented with protein-based chips[9–11].

Protease-specific protein chips. In principal, we envisage that the second type of protease chip — protease-specific protein chips — would incorporate arrays of molecules on solid supports or in nanowells that are designed to capture and assay specific proteases from complex biological samples. Despite the promise of such techniques, as shown in other systems[9–11], specific proof-of-concept papers that relate to proteases have yet to be published. However, protease-specific protein chips could be based on immobilized antibodies against different proteases or on protease-specific chemical reagents that have the ability to capture individual enzymes from complex mixtures. Retained proteases could be detected with antibodies or trypsin digested *in situ* and analysed by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF; BOX 2) and identified by PEPTIDE FINGERPRINTING (BOX 3) or by sequencing the tryptic peptides by mass spectrometry/mass spectrometry (MS/MS; BOX 3), which is also called tandem mass spectrometry.

PEPTIDE FINGERPRINTING
A mass spectrometric technique of protein identification that matches tryptic peptide masses of an unknown protein with those that are generated *in silico* for all the proteins in a database.

Box 2 | Mass spectrometry: a powerful tool for proteomics

Mass spectrometry measures the mass of proteins or peptides from analysis of the mass-to-charge (m/z) ratio. Mass spectrometers first ionize the sample, and then the ions are introduced into the mass analyser, which separates and detects the sample ions according to their mass. Fragmentation of the protein sample results in a collection of ions that have different masses, and the spectrometer measures the relative abundance of each ion, according to their m/z ratios, to obtain a spectrum of masses, hence the term mass spectrometry. Proteomic analyses typically use matrix-assisted laser desorption-ionization (MALDI) or electrospray ionization (ESI) sources followed by introduction into time-of-flight (TOF) or quadrupole (Q) mass analysers, which are described in more detail below.

Ionization sources

Matrix-assisted laser desorption-ionization (MALDI). Protein or peptide samples on solid matrices are ionized by a pulsed laser, entrained in the resulting ionized gas plume, and introduced into the mass analyser.

Electrospray ionization (ESI). Protein or peptide samples are passed through a fine needle to which a voltage is applied, which results in a fine spray of sample-containing droplets. Samples are delivered to the mass analyser after the breakup and evaporation of the droplets, which releases the protein or peptide samples to the gas phase. A low-flow-rate liquid chromatography system can be coupled to the needle to allow for protein fractionation of the samples before mass analysis.

Mass analysers

Time-of-flight (TOF). Time-of-flight analysers separate ions on the basis of their flight times over a known distance. The lower the mass of the ion, the greater the velocity and hence the shorter the flight time. Travel time from the ion source to the detector is transformed into the m/z ratio, from which the mass of the ionized sample can be calculated with extreme accuracy.

Quadrupole (Q) mass filter. A quadrupole mass filter consists of four parallel rods through which direct current and radio frequency electric fields are applied to sort the introduced ions. For each combination of voltages and frequencies, only ions with a specific m/z ratio pass undeflected through the quadrupole mass filter. Precise stepping of these settings therefore allows the quadrupole to be used as a mass analyser to scan for ions over a large m/z range.

Sensitivity of mass spectrometry, either of captured proteases on chips or from solution (as described later), however, is still an issue for obtaining complete coverage of the protease repertoire. To identify proteases that are involved in protein catabolism, this should not pose a problem, as the cell typically achieves this function by releasing relatively large quantities of proteases. For instance, this would apply to cathepsins, which are released into intracellular vacuoles or specific lysosomes that contain phagocytosed extracellular material that is destined for complete degradation, and to metalloproteinases such as the MMPs[2,12] and the related ADAMTS ('a disintegrin and metalloproteinase domain with thrombospondin modules') proteases[13], which are secreted outside the cell with the potential to degrade extracellular matrix proteins in pathological tissue destruction. However, differential concentration and separation by LC will be needed to enhance the signature of low abundance proteases. For example, the highly selective and precise processing of intracellular signalling and regulatory molecules, or extracellular bioactive molecules, can be achieved by proteases that are expressed at low levels. These are under the tight control that is exerted by highly regulated transcription and expression pathways. Indeed, it is expected that the proteases that are expressed at low levels — which initiate, modulate or terminate information cascades — are the very enzymes for which it is desirable to obtain this type of information.

A second, very important limitation of these techniques is that they focus on recording variations in protease-expression levels rather than on analysing their activity. This limitation, which is shared by all expression-based chips, is especially relevant in the case of proteases that are subject to several mechanisms of post-translational regulation, including zymogen activation, autocatalytic shedding of substrate- or inhibitor-binding domains, and inhibition.

Activity profiling. To advance from expression degradomics to functional degradomics, a third type of analysis — activity profiling — is required, which will lead to the development of protease-activity chips. Activity profiling is used to broadly survey the proteolytic activity of the different enzymes in complex samples. Importantly, it distinguishes active enzymes from their inactive-precursor or inhibitor-bound forms. The first attempts at functional degradomics have been directed at profiling serine- and cysteine-protease activities in crude protein samples[14–16]. As the catalytic mechanism of cysteine and serine proteases uses a covalently bound acyl intermediate, these approaches are based on the reactivity of these proteases towards tagged chemical probes in which the structure of a general inhibitor serves as a scaffold to which only active proteases can covalently bind. Active site-bound proteases are then identified by molecular weight and by fluorescent tag colour after electrophoretic migration past a fluorescence detector, or on western blot transfer to membranes[16]. An alternative approach, now in development in our laboratories, involves sequencing active proteases that have been captured by inhibitors, or other biotin-tagged affinity probes, and immobilized on affinity columns (FIG. 2a). After elution from avidin LC columns the captured proteases can then be identified by tandem mass spectrometry (BOX 3). This LC–MS/MS approach will be particularly useful for low abundance proteases, which can now be injected into the mass spectrometer in a more concentrated form.

Protease-activity protein chips. The next logical improvement of the protease-specific protein chip and protease-inhibitor LC approaches would be the immobilization of protease-specific inhibitors — either natural protein or chemical reagents — on chips to capture active proteases. Captured proteases could be identified in array readers using fluorescent-labelled antibodies or by quenching fluorescent-labelled capture probes, or, more directly and specifically, using MALDI-TOF or MALDI-quadrupole-TOF (MALDI-Q-TOF; BOX 2) tandem mass spectrometry after trypsin digestion *in situ*.

Chemical proteomics and imaging. The development of other types of tagged irreversible inhibitors, which are reactive against other classes of proteases, should provide an invaluable tool for the rapid detection, identification,

ADAM AND ADAMTS
Cell surface (ADAM) or secreted (ADAMTS) metalloproteinases that are related to MMPs, but that have a different multidomain structure, which includes a cysteine-rich disintegrin domain and thrombospondin modules.

Box 3 | Protein identification by mass spectrometry

Proteins can be identified by mass spectrometers by using two main methods — peptide fingerprinting and sequencing. Matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF; BOX 2) mass spectrometry (MS) precisely measures the masses of the tryptic fragments of a protein. To identify the protein, this peptide fingerprint is compared with predicted tryptic peptides generated *in silico* from all the proteins in a database. If protein identity cannot be established by peptide fingerprinting (because of ambiguity), the peptides can then be sequenced by MS/MS, which is also called tandem mass spectrometry. Two separate stages of mass analysis are carried out on tandem mass spectrometers using two integrated mass analysers (BOX 2), which are coupled together in the same machine. First, tryptic-digest samples are ionized either by MALDI or by nanoelectrospray ionization (BOX 2). The first mass spectrometer is used to select for peptide ions of known mass, the selected ions are then fragmented, and the second mass analyser measures the masses of the daughter ions from which the sequence can be derived.

Tandem mass spectrometry can be carried out using an ion-trap mass spectrometer, in which all tryptic peptide ions, except the ions of a selected mass, are expelled from the trap. Increasing the energy of the trap then fragments this peptide, with the mass detector measuring the masses of these fragments for peptide identification. Other examples of tandem mass spectrometers include triple quadrupole mass spectrometers — in which the second quadrupole is an argon gas collision cell — and the quadrupole-TOF mass spectrometer. The latter state-of-the-art instrument uses a quadrupole mass filter to select for ions of a desired mass, which are then fragmented in a gas collision cell, with the masses of these daughter fragments then being measured extremely accurately by a TOF mass analyser to derive their sequence.

isolation and even imaging of active enzymes that are present in cell and tissue proteomes. In what has been termed 'chemical proteomics'[16], the use of protease-specific or broad-spectrum protease inhibitors to achieve chemical knockouts of proteases promises to be a powerful new approach to uncover the biological roles of proteases in culture or *in vivo*. Moreover, fluorescent dyes coupled to inhibitors can be used to localize the cellular and tissue distribution of active proteases in histological sections[16]. In mouse models — and in humans in the future — the use of specialized labels coupled to protease inhibitors can be used for INTRAVITAL IMAGING of active proteases in tissues and in pathological lesions such as tumours. For example, using fluorescence-reflectance imaging, fluorescence-mediated tomography and near-infrared fluorochromes, active MMP-2 (gelatinase A), cathepsins and caspases have been localized in living tissues[17–19]. Other techniques for intravital imaging of active proteases, such as positron-emission tomography and multiphoton confocal microscopy, will also be possible[19], some of which are based on using quenched fluorescent synthetic substrates that generate a signal only on cleavage.

There are significant challenges to these approaches. The metalloproteinases do not have a covalent-bound substrate transition state, and there are no known covalent inhibitors that can be chemically developed for denaturing approaches. Innovative native strategies will be required for these enzymes — potentially based on biotin-tagged hydroxamate peptidic inhibitors — which could be used to separate these proteases from mixtures under mild native conditions. Elution and identification by LC–MS/MS would then follow. Although specificity might be built into these probes for proteases with unique inhibition or cleavage properties, if we consider the robustness of proteolytic systems, with their multiple overlapping substrate specificities, it is unlikely that it will be possible to fully cover the proteolytic landscape with protease-specific inhibitor probes. A potential variant of this might be to use broad-spectrum-family- or subfamily-specific inhibitors to profile proteases on a system-wide basis, and to use LC–MS/MS to then identify the inhibited proteases. However, this approach is not necessarily quantitative, so the relative abundance of individual proteases would not be reliably determined.

Protease isotope-coded affinity tag labelling. A recent technique that does provide quantitative data is based on protein labelling by ISOTOPE-CODED AFFINITY TAGS (ICAT)[20]. Potentially, protease ICAT will prove useful for mass-spectrometric identification and quantitation of proteases that are differentially expressed by cells or tissues in two distinct physiological or pathological conditions, such as normal and arthritic cartilage, or tumours (FIG. 2b). In this approach, an isotope-coded affinity tag — with either eight deuteriums (d8) or eight hydrogens (d0) and a biotin moiety — is used to label reactive cysteines in reduced protein samples or, potentially, to label the active-site residue in cysteine proteases. Identical proteins in two samples can be quantitated and compared by labelling one sample with the 'heavy' tag (d8) and the other sample with the d0 tag. After tryptic digestion, avidin-bound peptides are eluted from a nanoscale capillary affinity column, with identical peptides co-eluting and entering the mass spectrometer simultaneously. The relative abundance of the labelled proteases in the two samples is quantitated by mass spectrometric measurement of the 8-Da mass difference that is imparted by the heavy tag. Tandem mass spectrometry is then used to sequence and identify the eluted proteins (FIG. 2b). How can this technique be specifically adapted for proteases? The specificity of ICAT analyses might be improved by substituting covalent protease inhibitors for the cysteine tag to enable active proteases in biological samples to be quantitated by targeted ICAT. Hence, isotope-coded inhibitors that are based on irreversible inhibitors could be used for cysteine and serine proteases, for example.

Substrate chips. As the net cleavage of a particular substrate in a cell or tissue determines the biological outcome, valuable information can be gained from analysis of the net proteolytic potential of the entire functional degradome towards a particular substrate, without discriminating between the individual contributions of different proteases. At a system level, it is as important to quantitate whether a particular substrate is cleaved at a particular location, time and rate, as it is to identify the individual proteases that are present. So, a new fourth class of chip that we call 'substrate chips', which are based on protein, protein-analogue or peptide-cleavage and the removal of fluorescent end-labelled tags, can be envisaged (FIG. 2c). The development of protease-substrate chips is highly feasible, with the technology being adapted from similar applications for other systems. For example, protein-kinase chips have been developed to assess the ability of kinases to phosphorylate immobilized

INTRAVITAL IMAGING
Visualization of biological processes in intact animals or organ systems.

ISOTOPE CODED AFFINITY TAGS (ICAT). ICAT probes have different masses, but are chemically identical. They incorporate a reactive cysteine, a biotin moiety, and eight deuteriums in place of eight hydrogens, and they are used to specifically label, by mass-difference, identical proteins in two separate samples for the identification and semiquantitative comparison of abundance.

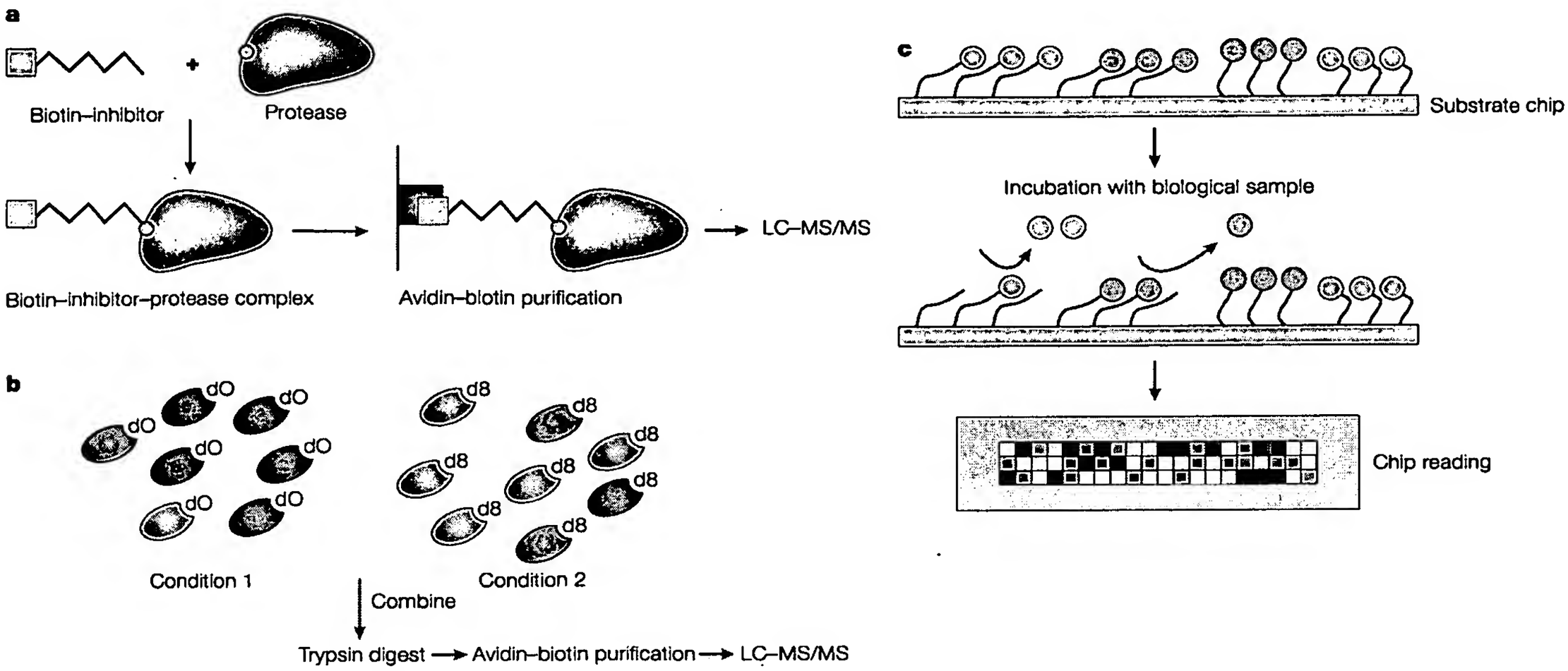


Figure 2 | **Degradomics approaches: activity profiling. a** | Protease-specific inhibitors that are coupled to biotin are used to probe for active proteases in complex biological mixtures. After affinity purification on avidin columns the captured proteases can be identified after trypsinization by liquid chromatography–mass spectrometry/mass spectrometry (LC–MS/MS). Alternatively, inhibitors that are arrayed on chips can capture active proteases, which can be identified by matrix-assisted laser desorption-ionization quadrupole time-of-flight (MALDI-Q-TOF) mass spectrometry after *in situ* trypsinization on the chip. **b** | Targeted isotope-coded affinity tag (TICAT) identification of proteases that are expressed by cells or tissues under two different conditions, such as normal and inflamed tissue, tumour and peritumour stroma tissue, or cells in culture that are treated with different growth factors. The active-site cysteine in the cysteine proteases, or elsewhere in other proteases, is reduced and alkylated using deuterium-8 (d8)-labelled tags and is compared with samples in the other tissue conditions that are similarly labelled, but not deuterium tagged (d0). The different proteases in the samples are represented by different colours in the figure. After mixing the two samples, the proteins are digested with trypsin and the biotin-tagged peptides are purified on avidin columns. Sample analysis is therefore greatly simplified by reducing the number of peptides for separation and analysis. For co-eluting, and therefore identical, peptides, the relative abundance of the d0- and d8-labelled peptides is quantified by nanoscale LC–MS, with nanoscale capillary LC–MS/MS used to identify the protease by peptide molecular weight and amino-acid sequence. **c** | Substrate chips could analyse net proteolytic activity towards specific scissile bonds that are present in synthetic peptides, protein substrates or protein analogues that are arrayed on chips. Substrates are fluorescent end-labelled and incubated with the samples, or smaller combinatorial peptides can be labelled by quenched fluorescent flurochromes. Loss or gain of signal, respectively, at the corresponding array position measures activity toward specific substrates. Although this approach could be successful using a single fluorescent dye for all the arrayed substrates, as new dyes become available, higher density chips will be possible, with the added potential of multiple labelling of proteins, for example, labelling individual subunits.

substrates[21,22]. A system-wide approach should prove valuable for the molecular diagnosis of patients and the molecular analysis of both the catabolic and processing actions of proteases. For example, the degradation of extracellular matrix proteins might be important for the diagnosis of cancer grade, or for the severity of arthritis and other inflammatory diseases. Determination of the tissue or cellular propensity for the proteolytic activation of bioactive molecules, such as Fas ligand[23], α-defensin[24], interleukin (IL)-8 and tumour-necrosis factor-α[12], or the inactivation of factors such as the chemokine immune mediators[8,25] by MMPs, would provide valuable information on the role of proteases and proteolytically processed factors in the initiation, development and resolution of many types of pathology.

What benefits will be gained from these studies? Degradomics analysis of the protease systems that are active in biological samples on a system-wide scale using activity profiling and substrate chips will provide a new level of information that is not available at present. System-wide analysis of whole families of proteases will provide data on both the synergy and functional redundancy of different proteases, and on their relative roles in different tissues or diseases. This should provide insights into how a cell or tissue responds in terms of initiating protease action in different physiological and pathological processes. So, different tissues, or the same tissue in different diseases, might achieve similar net substrate cleavage using different sets of proteases. Understanding this will prove extremely valuable in the design of therapeutic strategies to modulate proteolytic activity *in vivo*. For example, a system-wide analysis that identifies one or a few proteases that are responsible for processing a target substrate would logically direct drug development along highly specific lines to avoid side effects.

Alternatively, these analyses might point to large numbers of proteases that as a group are pathologically crucial, but that are individually redundant. In these instances, such as in the metastatic spread of cancer, it would be more appropriate to develop and administer a broad-spectrum inhibitor, provided that any side effects are manageable.

The proposed approaches to degradomics based on protease chips still have many of the limitations that are common to any approach that is derived from protein analysis by array technology (TABLE 2). These include problems with surface chemistries, nonspecific adsorptions, array instability owing to loss of native protein conformation, and the lack of robust and quantitative detection methods[26]. Despite these limitations, all of the proposed protease chips that have been discussed have either already been tested experimentally or can be adapted from similar applications in other systems. These preliminary results, as well as rapid advances in this field[16,27,28], show great promise for the widespread application of these approaches to degradomics.

Proteases in search of substrates

In addition to strategies that are aimed at profiling the protease complement of each organism or tissue at a given time, there is also a requirement for new approaches to examine the other side of the protease world — the identity of the substrates that are targeted by proteases. Several strategies have proven useful to address this problem over the past decade, and new advances have been reported recently. These approaches can be classified into two broad categories — genetics-based and proteomics-based — although there is significant overlap between them.

There are several recent studies in which *in vivo* protease substrates have been identified genetically[2,24,29,30]. These studies are based on the detection of non-processed substrates that have accumulated in the tissues or body fluids of mutant mice that are deficient in specific proteases. Complementary genetic approaches aimed at generating mutant mice that express noncleavable target substrates have also been reported[31]. However, the generalization of model-system-based genetic approaches for substrate identification might be hampered by the frequent occurrence, in all organisms, of robust proteolytic systems with redundant and compensatory enzymes. Moreover, directly ascribing a specific role to a protease can be problematic if proteolytic cascades are initiated by a target protease that has downstream effector proteases, which, in turn, are responsible for the cleavage of the substrate.

Nevertheless, knockout technology has been successfully used to determine the physiological roles of many proteases. For example, the use of gene-disrupted animals has provided interesting information on the role of caspases in important developmental and homeostatic programmes, including cardiac and neuronal development[32]. *Caspase-1*- and *caspase-11*-null mice do not process and activate IL-1β, IL-1α, IL-18 and γ-interferon[33], which indicates a role for these enzymes in cytokine activation. Similarly, gene-targeting of lysosomal cysteine proteases has confirmed the relevance of cathepsin S and cathepsin L in antigen presentation[34–36], and has uncovered new functions for some of these proteases, including a role for cathepsin L in hair-follicle morphogenesis and cycling[37]. Likewise, analysis of cathepsin-C-deficient mice has shown an *in vivo* function for this enzyme in the activation of several serine proteases, such as cathepsin G, granzyme A and granzyme B, neutrophil elastase, and chymase[38], and, from the analysis of *cathepsin-B*-knockout mice[39], this protease has been implicated in intracellular trypsinogen activation and the onset of acute pancreatitis.

Current methods for substrate identification. The biochemical identification of substrates has been typically carried out in a serial manner using defined substrates; however, the identification of the key *in vivo* substrates for many proteases will require more general proteomics-based methods. With the advent of mass spectrometry, cleaved substrates can be identified on a much larger scale by comparative analysis after the separation — using two-dimensional gel electrophoresis — of complex cell- or tissue-protein extracts that have been exposed to proteases, but with the limitations that were discussed earlier. Other global screening methods include phage-displayed peptide libraries, combinatorial fluorogenic substrate libraries or positional scanning libraries coupled with highly selective affinity labels[40–43]. Peptide-library screens, combined with database searching, provide a means for the rapid identification of potential protein substrates with consensus cleavage sites. Combinatorial peptide libraries, either end-labelled or quench-fluorescent labelled and arrayed on specialized substrate chips, might provide the ability to profile the amino-acid and scissile-bond preferences of individual proteases. However, considerable technological hurdles have to be overcome in the synthesis of combinatorial peptide libraries on chips *in situ*. The preference of active sites for particular amino-acid sequences can also be determined by combinatorial peptide-based inhibitor libraries[42]. Compared with peptide-substrate libraries, a particular advantage of this approach is that specific inhibitors are also identified that, if they are tagged with fluorescent dyes, can be used for activity profiling and histological[16] or intravital imaging. Inhibitors that are coupled to biotin can also be used for affinity purification and LC–MS/MS identification of proteases in biological samples, as discussed above. However, the identification of preferred scissile bonds does not identify substrates, and identification of substrates *in vitro* does not necessarily mean that the substrate is biologically relevant *in vivo*. Temporal and spatial localization of a protease with its substrate is essential for substrate recognition and cleavage *in vivo*. Structural conformation and exposure of potential scissile bonds to the protease are clearly also important for successful catalysis. Substrate-binding sites — called EXOSITES — which are located on specialized domains or modules outside the catalytic domain, and which can be crucial for substrate recognition and cleavage[12], are also neglected by these methods.

EXOSITE
A substrate-binding site that lies outside the active-site cleft of a protease and that is located on specialized substrate-binding modules/domains. Exosites generate diversity in substrate specificity and produce increased rates of cleavage.

EXOSITE SCANNING
On the basis of the hypothesis that proteins that bind protease exosites might be substrates, this is a technique for substrate identification that uses recombinant exosite domains as bait (see REF. 8).

INACTIVE CATALYTIC DOMAIN CAPTURE
(ICDC). A technique used for the identification of protease substrates. This uses mutated, proteolytically inactive catalytic domains to capture potential substrates without cleavage and subsequent release.

Exploiting exosites. Exosites are essential for the efficient cleavage of complex substrates, such as native triple-helical collagens by MMPs[12]. Exosites have also been found to be crucial for the binding and presentation of simpler molecules — such as chemokines — to the active site of MMPs for cleavage[8,25]. 'EXOSITE SCANNING' is a recent substrate-screening method that recognizes the importance of exosites in proteolytic function. The initial strategy that was reported was a new application of the yeast two-hybrid system, which used these ancillary protease exosite domains as bait to screen cDNA libraries for potential binding partners[8] (FIG. 3). The strategy is based on the hypothesis that proteins that bind exosites might be substrates of the protease. Exosite scanning with disulphide-bond-containing domains of extracellular proteases as bait for extracellular substrates was first shown with the identification of unexpected cytokine substrates for well-known proteases, such as for MMP-2, and the discovery of potential new roles for MMPs in the regulation of essential processes like inflammation[8] and HIV infection[25]. This approach should be very useful in the characterization of the growing number of multidomain proteases, including members of the ADAM, ADAMTS and membrane type serine proteinase (MTSP) families. These proteases contain many diverse ancillary domains (mostly of unknown function), and their specific substrates remain largely unknown[13,44].

It cannot be expected that all exosite domains will fold properly in yeast host cells, nor can it be expected that all substrates will be expressed and form functional complexes with exosite domains in the yeast nucleus, particularly if post-translational modifications are important in substrate recognition. However, protein substrates that are naturally present at low levels *in vivo* — such as growth factors and some cell-surface molecules — can be potentially identified by genetic methods such as this, more readily than by proteomic screens, in which abundance sets sensitivity limits. This strategy has also been adopted for proteomic screens of complex protein mixtures in which immobilized exosite-domain proteins are used to bind potential substrates that are identified by nanoscale capillary LC–MS/MS (FIG. 3). For those proteins that are not expressed effectively in yeast cells, this offers an alternative approach and is actively in development in our laboratories.

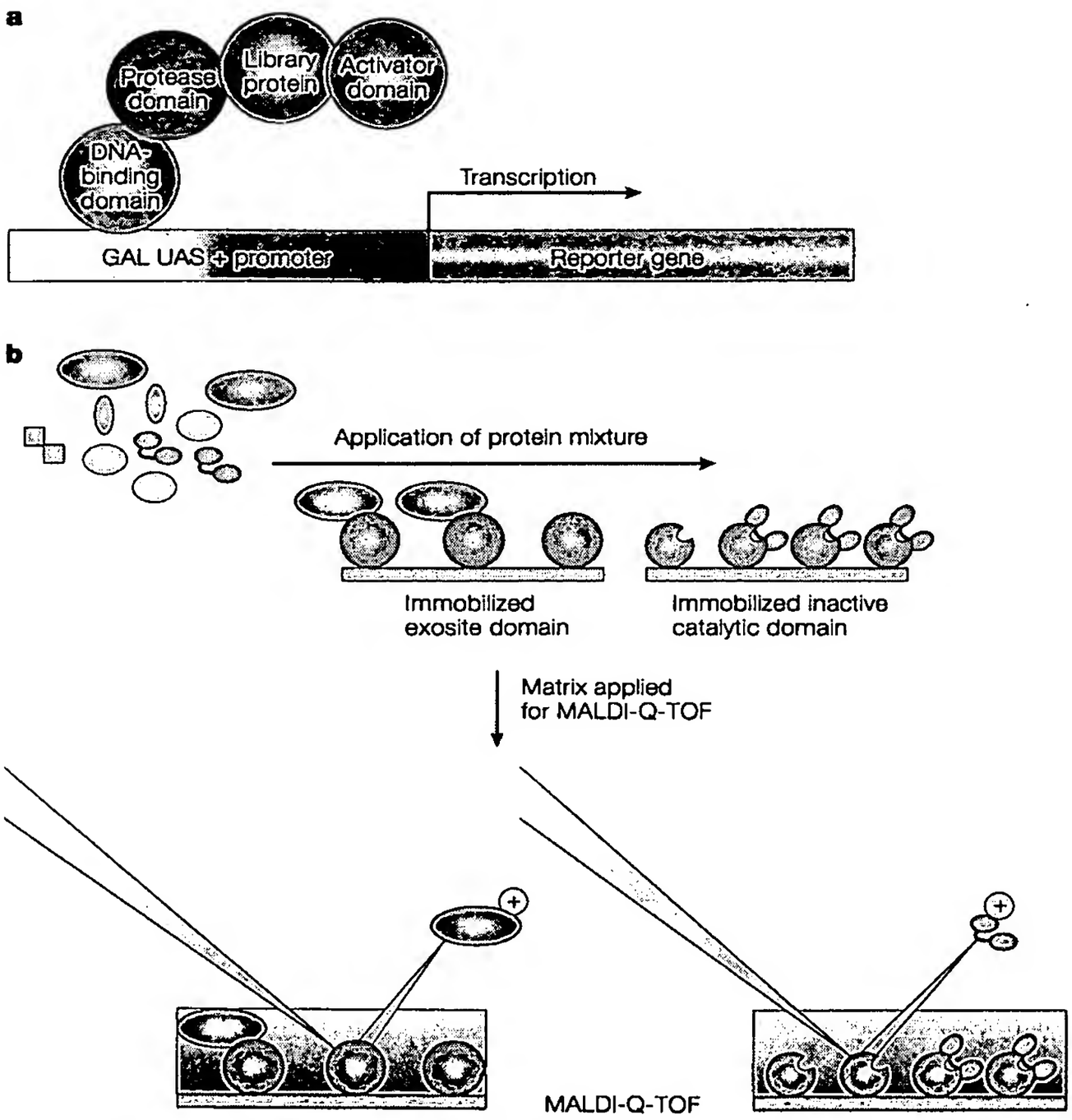


Figure 3 | **Exosite scanning and inactive catalytic domain capture (ICDC): substrate scanning by yeast two-hybrid screens and immobilization techniques. a** | In exosite scanning and ICDC, the yeast two-hybrid baits are the protease exosite domain or inactive catalytic domain, respectively (labelled 'Protease domain' in the figure), which are expressed in yeast as a fusion protein with the GAL4 DNA-binding domain ('DNA-binding domain'). Co-transfected into the cells is a second plasmid that expresses the GAL4 activation domain ('Activator domain') to which cDNA from a library is fused ('Library protein'). On interaction of the protease exosite with a library protein in the yeast nucleus, the DNA-binding and -activation domains are brought into proximity, which activates the reporter gene. In yeast two-hybrid ICDC, substrates that are bound to the mutated active site of the protease catalytic domain cannot be cleaved, which means that a positive transcription signal can be initiated. Interactor proteins that are identified are subsequently assessed biochemically for binding and cleavage by the protease. **b** | Exosite scanning and ICDC can also be used in proteomics screens, in which the exosite domain or inactive catalytic domain are immobilized on columns, solid matrices or chips to capture potential protein substrates. The column-bound substrate proteins can be eluted, trypsin digested and identified by tandem mass spectrometry or, for solid matrices and chips, digested *in situ* and identified directly by MALDI-Q-TOF mass spectrometry. MALDI-Q-TOF, matrix-assisted laser desorption-ionization quadrupole time-of-flight; UAS, upstream-activating sequence.

Inactive catalytic domains as substrate baits. Although our recent data indicate that exosites are crucial for the efficient turnover of certain substrates, or in some cases select against the cleavage of exposed peptide sequences that are rapidly cleaved when present in synthetic peptides, many substrates do not require exosite assistance. Therefore, in what we term 'INACTIVE CATALYTIC DOMAIN CAPTURE' (ICDC), we are also using inactive catalytic-domain mutants as baits in yeast two-hybrid screens and as immobilized native-protein baits for genetic and proteomic profiling, respectively, to identify new substrates (FIG. 3). In the absence of cleavage, proteins that are bound to the active site can then be identified. Known MMP substrates — such as laminin, galectin-3 and collagen — and some previously unknown cytokine substrates, have been cloned using genetic ICDC of membrane type 1-MMP (MT1-MMP) in yeast two-hybrid screens. Similarly, MMP-2 ICDC proteomic screens of cell and tissue extracts have also confirmed the utility of this technique.

Changing views of proteases

Approaches such as those that are described above and new technologies that are now being developed in our laboratories and by others, which use variants of ICAT and innovative techniques for sample preparation before

tandem mass spectrometry, promise to uncover many new and unexpected substrates and functions for proteases. Even now, the identification of new protease substrates is redefining our outlook on the *in vivo* roles of certain proteases. Over the past 10 years, the identification of an enormous array of cytoplasmic and nuclear-protein substrates of the intracellular cysteine caspases[33] has provided convincing evidence for the exquisite degree of control that can be executed by proteolytic processing in pivotal cellular processes, such as apoptosis. In addition to the activation of downstream caspases, the highly selective scissile-bond specificity that is shown by caspases is manifested by limited proteolysis and constitutive activation of several kinases, including several protein kinase C isoforms and the mitogen-activated protein kinase/extracellular-signal regulated kinase (MAPK/ERK) kinase kinase MEKK-1. Other targets, however, are inactivated by caspase processing, including p53 and retinoblastoma susceptibility protein (pRb), focal adhesion kinase (FAK), the transcription factors nuclear factor-κB and signal transducer and activator of transcription-1 (STAT-1), and the cytoplasmic proteins β-catenin and the anaphase-promoting complex (APC) gene product[33,45].

As another example of a change in the view of proteases, the MMPs have traditionally been thought to be the effectors of extracellular matrix catabolism in normal remodelling and pathological tissue destruction. This view was modified about 10–15 years ago, when the influence of MMPs on cell behaviour was recognized and ascribed to the disruption of the pericellular matrix and cognate integrin contacts that results in alterations in cellular signalling[2]. MMP production by cancer cells was also thought to lead to tumour-cell invasion and metastasis. However, it is now widely recognized that most MMP expression in cancer is from the reactive tissue stroma. Indeed, a recent example of a new role for MMPs in cancer is where selective MMP production by cancer cells drives selection for tumour-cell resistance to apoptosis[23]. This example is indicative of the marked revision in our concepts of the roles of MMPs in cancer and other pathologies. Moreover, the growing awareness that MMPs cleave a wide range of bioactive substrates (see online link to Bioactive MMP substrates) *in vivo*[8,24,29]; the relatively minor phenotypic alterations in normal matrix turnover in most MMP-knockout mice; the general absence of effects on normal connective-tissue remodelling during long-term exposure to synthetic MMP inhibitors in animal models and in clinical trials with MMP-inhibitors; and the existence of efficient intracellular pathways of extracellular matrix turnover[46] indicate that MMPs are not important for normal matrix remodelling as generally assumed. Indeed, the general lack of *in vivo* evidence reported for extracellular-matrix-protein cleavage by MMPs, even in pathology, indicates other important roles for these proteases. We suggest that bioactive substrate processing by MMPs, with its resulting regulatory effects on the cellular responses of the tissue and host defence cells, in physiological and pathological processes will prove to be more important overall than their potential catabolic roles. Hence, as new substrates are described for all proteases, it is anticipated that the biological roles for several protease families will be viewed quite differently in the years ahead, with many showing new potential as drug targets to treat disease.

Outlook

The completed genome projects have provided a good overview of the composition and organization of the proteolytic machinery that is used by different organisms. However, not all proteases have been discovered yet, nor have their *in vivo* substrates and functions been identified. Indeed, many human protease genes that have been predicted from genome computer searches remain to be cloned and the proteolytic activity of their gene products confirmed. An important challenge for the future is to solve the structure of, and to assign physiological and pathological functions to, the proteins that are encoded by these new protease genes. Other proteins that have significant identity to proteases but that lack catalytic residues (such as several members of the ADAMs family) — which we propose might function as natural highly specific antagonists by binding substrates without cleavage, thereby masking scissile bonds — will also need to be characterized to determine whether they contribute to the fine-tuning of protease activity *in vivo*. Furthermore, new structural designs for proteases that would not be detected by homology-based scanning or cloning methods might remain hidden in our genome[3].

Genomic-based approaches to degradomics will be essential for determining the basis of the increasing number of genetic diseases that are recognized as being caused by mutations in protease loci. These genomic approaches will also be essential for the identification of putative single-nucleotide polymorphisms in protease genes that might confer increased susceptibility, or resistance, to complex diseases[47]. In parallel, the development of strategies for intravital imaging of active proteases, and the generation of new animal models with a gain or loss of protease function, should contribute to the evaluation of the role of individual proteases or protease families in health and disease. Chemical proteomics will be a powerful complement to determine the roles of protease families as a proteolytic system or as specific proteases individually in culture or *in vivo*. It will also be essential to define the regulatory and functional intersections between proteolytic systems and other important networks, such as those involved in signalling, and to identify new protease circuits that function in distinct cells and biological processes.

No single screening technology will provide complete genomic or proteomic coverage to identify the entire protease degradome or the substrate degradome of a protease, and each technology has its own specific advantages and niche. However, emerging technologies for high-throughput assays, including protease chips, will greatly assist in generating the necessary data for profiling proteases in proteomes and for the identification of substrates and inhibitors, thereby defining new therapeutic targets. Similarly, determining the three-dimensional structures of proteases by new high-throughput

X-ray crystallographic approaches[48], or by using a combination of homology modelling and *ab initio* predictions that aim to predict the three-dimensional structure of a protein from its amino-acid sequence[49], will be valuable resources for rational inhibitor design. Again, there are many challenges ahead, especially in the case of membrane proteases that are responsible for processes such as regulated intramembrane proteolysis[50]. By combining our knowledge at present with new and creative strategies, degradomics promises to provide answers to questions about the expression, control and *in vivo* roles of proteases — a large group of enzymes that profoundly influence cell behaviour, survival and death in all living organisms with precision.

1. Barrett, A. J. *et al.* (eds) *Handbook of Proteolytic Enzymes* (Academic, London, 1998).
A 'must-have' volume for any protease lab that comprehensively lists and describes numerous proteases.
2. Sternlicht, M. D. & Werb, Z. How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **17**, 463–516 (2001).
3. Urban, S., Lee, J. R. & Freeman, M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**, 173–182 (2001).
4. Balbin, M. *et al.* Identification and enzymatic characterization of two diverging counterparts of human interstitial collagenase (MMP-1) expressed at sites of embryo implantation. *J. Biol. Chem.* **276**, 10253–10262 (2001).
5. Deussing, J. *et al.* Identification and characterization of a dense cluster of placenta-specific cysteine and peptidase genes and related genes on mouse chromosome 13. *Genomics* **79**, 225–240 (2002).
6. Yousef, G. M. & Diamandis, E. P. The new human tissue kallikrein gene family: structure, function, and association to disease. *Endocrine Rev.* **22**, 184–204 (2001).
7. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
8. McQuibban, G. A. *et al.* Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3. *Science* **289**, 1202–1206 (2000).
The first description of exosite scanning and the use of the yeast two-hybrid system to identify new extracellular protease substrates.
9. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
This study shows the usefulness of protein-based microarray chips for functional analyses.
10. De Wildt, R. M. T., Mundy, C. R., Gorick, B. D. & Tomlinson, I. M. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnol.* **18**, 989–994 (2000).
11. Fung, E. T., Thulasiraman, V., Weinberger, S. R. & Dalmasso, E. A. Protein biochips for differential profiling. *Curr. Opin. Biotechnol.* **12**, 65–69 (2001).
12. Overall, C. M. Matrix metalloproteinase substrate binding domains, modules and exosites: overview and experimental strategies. *Meth. Mol. Biol.* **151**, 79–120 (2001).
13. Cal, S. *et al.* Cloning, expression analysis and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains. *Gene* **283**, 49–62 (2002).
14. Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **96**, 14694–14699 (1999).
Together with references 15 and 16, this paper shows the feasibility of activity-based profiling of proteases that are present in complex mixtures.
15. Greenbaum, D., Medzihradszky, K. F., Burlingame, A. & Bogyo, M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569–581 (2000).
16. Greenbaum, D. *et al.* Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1**, 60–68 (2002).
A key paper in the development of active-site functional probes for cysteine proteases. This shows that BODIPY-dye-conjugated protease inhibitors can be used to specifically label and identify proteases in mixtures without prior separation. This also shows their use in histological localization, and in the identification of active-site peptide-bond specificity.
17. Weissleder, R., Tung, C. H., Mahmood, U. & Bogdanov, A. Jr. *In vivo* imaging of tumors with protease-activated near-infra-red florescent probes. *Nature Biotechnol.* **17**, 375–378 (1999).
18. Bremer, C., Tung, C. H. & Weissleder, R. *In vivo* molecular target assessment of matrix metalloproteinase inhibition. *Nature Med.* **7**, 743–748 (2001).
19. Weissleder, R. Scaling down imaging: Molecular mapping of cancer in mice. *Nature Rev. Cancer* **2**, 1–8 (2002).
A comprehensive review of the current status of various new and emerging techniques for intravital imaging of cancer proteases that use labelled substrate analogues or inhibitors.
20. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
One of several key papers from the Aebersold laboratory on a new mass-spectrometric proteomic technique that identifies and semiquantitatively measures differences in the abundance of the same protein in two samples.
21. Williams, D. M. & Cole, P. A. Kinase chips hit the proteomic era. *Trends Biochem. Sci.* **26**, 271–273 (2001).
22. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283–289 (2000).
23. Fingleton, B., Vargo, T., Crawford, H. C. & Matrisian, L. M. Matrilysin expression selects for cells with reduced sensitivity to apoptosis. *Neoplasia* **3**, 459–468 (2001).
24. Wilson, C. L. *et al.* Regulation of intestinal α-defensin activation by the metalloproteinase matrilysin in innate host defense. *Science* **286**, 113–117 (1999).
25. McQuibban, G. A. *et al.* Matrix metalloproteinase activity inactivates the CXC chemokine stromal cell-derived factor-1. *J. Biol. Chem.* **47**, 43503–43508 (2001).
26. Mitchell, P. A perspective on protein microarrays. *Nature Biotechnol.* **20**, 225–229 (2002).
Together with references 27 and 28, this paper presents the current status, recent developments and limitations of protein- and peptide-microarray protein chips.
27. Heng, Z. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
28. Houseman, B. T., Huh, J. H., Kron, S. J. & Mrksich, M. Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnol.* **20**, 270–274 (2002).
29. Liu, Z. *et al.* The serpin α1-proteinase inhibitor is a critical substrate for gelatinase B/MMP-9 *in vivo*. *Cell* **102**, 647–655 (2000).
30. Pendas, A. M. *et al.* Defective prelamin A processing and muscular and adipocyte alterations in Zmpste24 metalloproteinase-deficient mice. *Nature Genet.* **31**, 94–99 (2002).
31. Zhao, W., Byrne, M. H., Wang, Y. & Krane, S. M. Osteocyte and osteoblast apoptosis and excessive bone deposition accompany failure of collagenase cleavage of collagen. *J. Clin. Invest.* **106**, 941–949 (2000).
32. Ranger, A. M., Malynn, B. A. & Korsmeyer, S. J. Mouse models of cell death. *Nature Genet.* **28**, 113–118 (2001).
33. Earnshaw, W. C., Luis, M. M. & Kaufmann, S. H. Mammalian caspases: structure, activation, substrates and functions during apoptosis. *Annu. Rev. Biochem.* **68**, 383–424 (1999).
34. Nakagawa, T. *et al.* Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. *Science* **280**, 450–453 (1998).
35. Nakagawa, T. Y. *et al.* Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in *cathepsin S* null mice. *Immunity* **10**, 207–217 (1999).
36. Turk, V., Turk, B. & Turk, D. Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* **20**, 4629–4633 (2001).
37. Roth, W. *et al.* Cathepsin L deficiency as molecular defect of *furless*: hyperproliferation of keratinocytes and perturbation of hair follicle cycling. *FASEB J.* **14**, 2075–2086 (2000).
38. Pam, C. T. N. & Ley, T. J. Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B *in vivo*. *Proc. Natl Acad. Sci. USA* **96**, 8627–8632 (1999).
39. Halangk, W. *et al.* Role of cathepsin B in intracellular trypsinogen activation and the onset of acute pancreatitis. *J. Clin. Invest.* **106**, 773–781 (2000).
40. Deng, S. J. *et al.* Substrate specificity of human collagenase 3 assessed using a phage-displayed peptide library. *J. Biol. Chem.* **275**, 31422–31427 (2000).
41. Harris, J. L. *et al.* Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl Acad. Sci. USA* **97**, 7754–7759 (2000).
Together with reference 43, this paper describes the identification of preferred protease scissile bonds in iterative peptide-library-based approaches.
42. Nazif, T. & Bogyo, M. Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl Acad. Sci. USA* **98**, 2967–2972 (2001).
43. Turk, B. E., Huang, L. L., Piro, E. T. & Cantley, L. C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nature Biotechnol.* **19**, 661–667 (2001).
44. Hooper, J. D., Clements, J. A., Quigley, J. P. & Antalis, T. M. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *J. Biol. Chem.* **276**, 857–860 (2001).
45. Nicholson, D. W. & Thornberry, N. A. Caspases: killer proteases. *Trends Biochem. Sci.* **22**, 299–306 (1997).
46. Sodek, J. & Overall, C. M. in *The Biological Mechanisms of Tooth Eruption and Root Resorption* (ed. Davidovitch, Z.) 303–311 (EBSCO Media, Birmingham, Alabama, 1988).
47. Zhu, Y., Spitz, M. R., Lei, L., Mills, G. B. & Wu, X. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter enhances lung cancer susceptibility. *Cancer Res.* **61**, 7825–7829 (2001).
48. Thornton, J. Structural genomics takes off. *Trends Biochem. Sci.* **26**, 88–89 (2001).
49. Bonneau, R. & Baker, D. *Ab initio* protein structure prediction: progress and prospects. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 173–189 (2001).
50. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**, 391–398 (2000).

Acknowledgements

C.M.O. is supported by a Canada Research Chair in Metalloproteinase Biology, and by grants from the Canadian Institutes for Health Research, the National Cancer Institute of Canada, the Protein Engineering Network of Centers of Excellence and the Canadian Arthritis Network of Centers of Excellence. C.L-O. is supported by the Comision Interministerial de Ciencia y Tecnologia, Spain, and by the European Union.

Online links

DATABASES

The following terms in this article are linked online to:

Flybase: http://flybase.bio.indiana.edu/
Rhomboid

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
α-defensin | Fas ligand | galectin-3 | IL-1α | IL-1β | IL-8 | IL-18 | γ-interferon | MEKK-1 | MMP-2 | MT1-MMP | tumour-necrosis factor-α

Swiss-Prot: http://www.expasy.ch/
Caspase-1 | *caspase-11* | β-catenin | *cathepsin B* | cathepsin C | cathepsin G | cathepsin L | cathepsin S | chymase | FAK | GAL4 | granzyme A | granzyme B | neutrophil elastase | p53 | pRb | STAT-1

FURTHER READING

Bioactive MMP substrates: http://www.clip.ubc.ca/mmp.shtm
Chris Overall's laboratory: http://www.clip.ubc.ca
Encyclopedia of Life Sciences: http://www.els.net/
confocal microscopy | liquid chromatography
MEROPS version 5.7: http://www.merops.ac.uk

Access to this interactive links box is free online.